

Mind games

How not to mix politics and science.

Politicians are rarely elected solely for the soundness of their policies. As the armies of image consultants and pollsters will attest, people often vote with their hearts, and nothing matters more than following the pulse of that elusive prize: the voters' emotions.

But how do you find out what people are feeling? In a blaze of colour on the 11 November 'op-ed' (invited opinion) page of *The New York Times*, some scientists proclaimed that, based on analysis of brain-imaging data from just a handful of swing voters, they had divined what the rest of the undecided masses truly think about the upcoming US presidential elections. Apparently just asking them was simply not good enough.

So how did they uncover the innermost thoughts of their 20 subjects? The authors used functional magnetic resonance imaging (fMRI) to scan the subjects' brains while they viewed images of political candidates. This imaging technique can be used to measure changes in oxygenated blood and hence to infer changes in metabolic activity in different parts of the brain. Some parts of the brain reliably alter their activity under certain conditions, and scientists have used this fact, along with information drawn from other techniques in both humans and animals, to document which brain area is associated with which cognitive function. For example, greater activity in the insula is often reported when people experience disgust, whereas more activity in the amygdala is reported when people are anxious.

The authors of the op-ed up-ended this logic. For example, they observed more activity in the insula while their subjects looked at a picture of Democrat hopeful John Edwards and deduced that they must be disgusted by him, albeit unconsciously. Bad news for John Edwards — except that there is an inconvenient truth gumming up the deductions: increased activity in any brain area is rarely exclusive to any one function. That insula activity did not necessarily mean the subjects were disgusted. Insula activity has also been associated with drug craving, the taste of chocolate, pain and the quality of orgasm. Not necessarily such bad news after all.

The op-ed work has not been published in a peer-reviewed journal, and the article is self-evidently too insubstantial in scientific detail to assess the strength of either the methods or the data. A group of

cognitive neuroscientists was swift to object to its conclusions — which veer close to a modern-day phrenology — in a response to *The New York Times*.

The results described in the op-ed are apparently the claims of a commercial product posing as a scientific study. This is only partially transparent. Three of the authors list their affiliation with FKF Applied Research, a company based in Washington DC that is notorious for using similar brain-scan analysis to conclude which TV adverts aired during a major sporting event were most effective. In its own words, the company is a "business intelligence firm selling fMRI brain scan-based research to Fortune 500 companies".

More troubling for a mainstream newspaper that prides itself on its balanced reporting is the absence of declarations from three other authors. Rightly listed as affiliated to a neuroscience institute at the University of California, Los Angeles, one is also a co-founder of FKF Applied Research and all three, according to a previous publication, have benefited from funding from the company.

Articles on *The New York Times* op-ed pages are opinionated by definition, and shouldn't normally require peer review. But here, the paper's editors have instead published the results of (to put it mildly) questionable scientific research, disseminating this information to millions of their readers who may not have the background to recognize for themselves the absurdity of some of the authors' conclusions.

Although it is a gross disservice to science and indeed to politics, it is a great deal for the company. Scientific publication would have required the authors to divulge their data and qualify their assumptions — and some journals might even have required that they declare their financial interests. Whatever the motives, seducing *The New York Times*' editors with the allure of Technicolor brains lighting up with Hillary Clinton angst yielded no more or less than a multimedia advertisement for the company's product to millions of readers.

And does anyone need a \$3-million scanner to conclude that Hillary needs to work on her support from swing voters? ■

"The paper's editors have instead published the results of (to put it mildly) questionable scientific research."

Replicator review

Nature has implemented a peer-review policy for strong claims.

The transfer of a nucleus from a somatic cell of one organism into an enucleated germ cell of another for therapeutic or reproductive cloning is now well-established in many species, but has proved notoriously difficult to do in primates. Indeed some experts have concluded that it simply couldn't be done. Woo Suk

Hwang's high-profile paper reporting that it had worked in human cells turned out to be fraudulent, making the goal seem even more elusive.

Nature has now published what we expect to be the final word on whether nuclear transfer can work in a primate — a paper by J. A. Byrne *et al.* showing not only that it is possible to clone primate embryos by somatic-cell nuclear transfer but also that precious embryonic stem cells can be derived from the embryos (see page 497). If embryonic stem cells live up to their promise, the technology could be used to derive patient-tailored stem cells. Even if they don't, the ability to clone primate embryos means that researchers may be able

to create cell lines from patients with disorders such as Parkinson's disease and diabetes, and use them to screen drugs and examine cell pathology (see page 462).

Nature took the unusual step of soliciting an independent verification of the paper during the process of peer review (see D. S. Cram *et al.* *Nature* doi:10.1038/nature06456; 2007). This is the first time that *Nature* has obtained second-party replication ahead of publication. It should not be seen as reflecting a mistrust of scientists in the cloning field or scepticism about this particular research group. Rather, our actions fulfilled a statement in an Editorial (see *Nature* 439, 243; 2006) that was conceived in the aftermath of the Hwang affair: "Keeping in mind the principle that extraordinary claims require extraordinary proof, *Nature* may in rare cases demand it."

The decision was partly pragmatic — the corroborating experiment was straightforward and an expert was willing to do it on a timescale that would not delay publication of the paper. It was also based on history. Past fraud is no reason to submit an entire discipline to unusually tough review, but a history in stem-cell research of difficulties in replication of genuine claims carries some weight. The experiments are tough to do, there are many 'moving parts' in nuclear transfer and, given the importance of the conclusions of this work, it is desirable to be doubly robust in excluding the possibilities that embryos were a result of parthenogenesis, cell-line contamination or sample mishandling. (A blog discussion of peer-review challenges in stem-cell research and the peer-review reports for the paper by

Byrne *et al.* can be found at *Nature Reports Stem Cells*; www.nature.com/stemcells.)

So *Nature* asked a team led by Alan Trounson of Monash University in Victoria, Australia, an expert on human embryonic stem cells, to obtain tissue samples from the donor animals and stem-cell lines, and to test for the origins of the stem cells' nuclear and mitochondrial DNA by genotyping. In successful nuclear transfer, the mitochondrial DNA will come from the oocyte donor and the nuclear DNA from the somatic-cell nuclear donor. An Oregon veterinary surgeon, Theodore Hobbs, collected tissue from the animals, coded it, observed DNA preparation in the authors' lab, and then shipped the samples to a laboratory at University of Southern California, where Monash scientist David Cram conducted the analysis.

Nature will continue to evaluate the need for such validation on a case-by-case basis. Meanwhile, researchers should consider maintaining tissue samples, and trying to establish and fulfil requirements that may go beyond the routine, as Byrne *et al.* did for their cloning paper. In that way, remarkable scientific developments can remain as free as possible from unwarranted speculation and controversy. ■

"Researchers should consider maintaining tissue samples, and trying to establish and fulfil requirements that may go beyond the routine."

Better late than never

The decision to make the Leopoldina Germany's national academy of sciences is to be welcomed.

The road to a national academy for Germany has been a long and rocky one. Back in 1990, just months after reunification, then science minister Heinz Riesenhuber asked the Leopoldina — the learned society with the longest tradition in the German-language part of the world — to take on the role. But having just emerged from 40 years of communist quasi-isolation, the leaders of the 355-year-old academy declined to do so.

Seventeen years, two government changes and four science ministers later, Germany finally has a national academy. And it is the Leopoldina that will form its basis (see page 470).

When science minister Annette Schavan, a Christian Democrat like Riesenhuber, announced on 16 November that the Leopoldina should fill this role, it this time accepted the invitation.

Germany's seven regional scientific academies were surprised, and in some cases annoyed, by the decision. The country's *Länder* (states) have far-reaching autonomy in cultural affairs, which is why designating the national voice in science to a single body has been problematic in the past. As a result, Germany lacks a body that can represent the national opinion on science to the outside world and to German policy-makers in the way that the Royal Society does in Britain, or the National Academies in the United States.

Schavan's move therefore deserves applause. It resolves, by

satisfyingly non-bureaucratic means, an issue that had threatened to become buried forever between folders and beneath reports. That it was made by virtual fiat, with little public consultation, doesn't spoil the effect.

Germany will benefit from a national academy that can provide genuinely independent advice on scientific questions. The challenges posed to the federal government by such matters as global warming, genetics and demographic change demand such a mechanism.

The current, fragmented academy landscape doesn't match the needs of a large modern nation such as Germany. Some would have preferred to set up a new body (see *Nature* 443, 371–372; 2006), but this is not actually necessary. The Leopoldina, formed in 1652 and based in Halle in eastern Germany, demonstrated in communist times its independence and its moral and scientific integrity. It has all it takes to serve as a national academy, and it should take up its work with confidence.

Germany's regional academies should accept the invitation to assist it in its new task, particularly by adding their strengths in the humanities, where the Leopoldina lacks expertise (the academy will be responsible for *Wissenschaft*, the German word for science that actually embraces the humanities as well as hard science). If the outcome proves to be as solid and deeply rooted in science as is the Leopoldina's reputation, then this decision was worth the 17-year wait. ■

"Germany will benefit from a national academy that can provide genuinely independent advice on scientific questions."

RESEARCH HIGHLIGHTS

Sweet insight

Biol. Lett. doi:10.1098/rsbl.2007.0496 (2007)

Researchers in South Africa have overturned a long-held belief that hummingbirds and perching birds consume different nectars. Steven Johnson of the University of KwaZulu-Natal and Susan Nicolson of the University of Pretoria suggest that a more accurate distinction falls between specialist nectar-feeders and omnivorous species.

Flowers make either small amounts of concentrated sucrose-rich nectar or larger volumes of dilute liquid containing simpler sugars. Conventional wisdom held that flowers pollinated by sunbirds — perching birds that specialize in drinking nectar — make dilute nectar similar to that of blooms providing for generalist perching birds (such as the Cape weaver, pictured), and unlike the concentrated nectar of hummingbird-adapted flowers.

These authors report that this is not the case, a finding that alters ecologists' ideas about the evolutionary links between flowers and their avian pollinators.

Differences in the digestive enzymes of bird guts may explain why the types of sugars found in the two recipes are distinct, the researchers add.



S. JOHNSON

INORGANIC CHEMISTRY

Triple connection

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0707035104 (2007)

The synthesis of molecules with triple bonds between uranium and carbon atoms is possible, according to the findings of researchers based in America and China.

Lester Andrews of the University of Virginia in Charlottesville, Jun Li at Tsinghua University in Beijing and their colleagues reacted uranium atoms with halomethanes. They then used infrared spectroscopy to show that the complexes formed contained relatively strong uranium-carbon triple bonds. Changing the atoms close to those in the bond could 'tune' its strength, says Andrews.

But creating these quirky molecules under normal laboratory conditions is a long way off. This team needed lasers to generate the uranium atoms, plus a caesium iodide block held at a frosty 8 kelvin in an argon-rich atmosphere to produce the complexes.

ACOUSTICS

Nostril navigation

Phys. Rev. E **76**, 051902 (2007)

The sonar squeak of a bat is governed by the shape of its nose, according to physicists at Shandong University in Jinan, China.

Using X-ray tomography of the face of *Rhinolophus rouxi* (pictured right), a type of horseshoe bat, the researchers created a three-dimensional model of the structures around its nostrils, known the noseleaf. Qiao Zhuang and Rolf Müller programmed sonar waves to propagate through their model, noting where the waves resonated inside the nose and what

shape the waves formed when they left it.

Low-frequency waves bounced off structures that shaped them into a wide beam. Meanwhile, high-frequency waves resonated in a separate part of the noseleaf that focused the sonar signal. This could explain why switching frequency seems to help horseshoe bats shift from general navigation to homing in on prey.

EPIGENETICS

The single life

Science **318**, 1136-1140 (2007)

Although endowed with two versions of most genes, humans seem to rely on just one surprisingly often. Geneticists had thought that, with rare exceptions, when a gene is turned on both its maternal and paternal versions are expressed. But Andrew Chess and his colleagues at Harvard Medical School in Boston, Massachusetts, have screened almost 4,000 genes in B lymphocytes, a type of immune cell, and found that 371

of them sometimes break this rule.

The authors found that the choice of which version these cells use is random, but that every daughter cell follows its progenitor's decision. Further study will firm up how this effect might influence human traits.

MOLECULAR BIOLOGY

Starting block

Cancer Cell **12**, 432-444 (2007)

An arrangement that inactivates genes on one of the two X chromosomes in female mammalian cells is also involved in certain cancers.

In some inactivated X-chromosome genes the sites at which transcription would begin are wrapped in structures known as nucleosomes, making them inaccessible. According to Peter Jones and his colleagues at the University of Southern California in Los Angeles, nucleosomes also become bound in — and inactivate — promoter regions of several genes with a tumour-suppressor role.

Working with several cancer cell lines, Jones and his team removed methyl groups, which draw in nucleosomes, from DNA in the promoter region of a gene that is frequently turned off in colon cancer. Doing so freed three start sites and switched the gene into a non-cancerous 'on' mode.

ORGANIC CHEMISTRY

Sifting for bonds

J. Am. Chem. Soc. advance online publication doi:10.1021/ja074155j (2007)

Reactions that form bonds between two chemical groups are the basic tools of molecule-building. These have mostly been



R. MÜLLER

discovered by trial and error, but David Liu and his co-workers at Harvard University in Cambridge, Massachusetts, have devised a method for screening large numbers of reactant pairs in organic solvents to find out which join up under different conditions.

It involves tethering two different chemical groups to a single strand of DNA that is unique to each combination. If two groups react, the product remains stuck to the pair-specific DNA, even though the DNA's link to one of the groups is cut.

A chemical hook attached to one reactant group sticks to the surface of magnetic beads, which can then be separated. But the hooks only retain the DNA of pairs that have reacted — otherwise the earlier cleavage process cuts it free. DNA hybridization then reveals the identities of the product-forming pairs.

PLANT EVOLUTION

Taking after mother

Science **318**, 1134–1136 (2007)

Mother plants can transmit information about their environment to their offspring, thereby increasing their progeny's fitness.

Laura Galloway at the University of Virginia in Charlottesville and Julie Etterson at the University of Minnesota–Duluth studied *Campanulastrum americanum*, a small forest plant that grows both in sunny spots and in the shade of the forest canopy. Its seeds tend to sprout near the mother plant, within the same patch of sun or shade.

When the researchers grew plants in the opposite light environment from that of their mothers, the population grew at a slower rate. This suggests that maternal plants cue their offspring about the light environment they are likely to experience. Because this signal can change each generation, it is a flexible way of enhancing the success of offspring.



MATERIALS SCIENCE

Out of reach

Nanotoday **2**, 44–47 (2007)

Space elevators (depicted left) are a mainstay of science fiction. But advances in nanotechnology have spurred Nicola Pugno at Torino Polytechnic in Italy to consider whether shooting a payload into space up a 100,000-kilometre-long cable may one day be possible.

Previously, he worked out that the tensile strength of a huge carbon nanotube bundle would drop from 100 gigapascals to 80 gigapascals, were each nanotube to have one atom out of place. Now his calculations show that each rod would inevitably contain at least one defect four atoms long, making it less than 45 gigapascals strong and weaker than the 63 gigapascals required for a line of uniform thickness. Manufacturing constraints would force engineers to make a cable up to 600 times fatter at the height of its geosynchronous orbit.

CELL BIOLOGY

Size control

J. Cell Biol. doi: 10.1083/jcb.200708054 (2007)

Fission yeast regulates the size of its nucleus to stay in sync with the volume of its cytoplasm, researchers have found.

How cells control organelle size is a century-old question. Frank Neumann and Paul Nurse at the Rockefeller University in New York measured nuclear and cellular volumes in the fission yeast *Schizosaccharomyces pombe*. Using various mutants, they found similar nucleus-to-cell volume ratios over a 35-fold variation in cell size and a 16-fold difference in DNA content.

And from studies of cells with four nuclei, they suggest that the cytoplasmic environment that the nucleus resides in is an important regulator of its size. The molecular basis of this control remains unknown.

QUANTUM PHYSICS

Early jiggles

Phys. Rev. Lett. **99**, 201301 (2007)

A line of ions could provide clues about the early Universe. Ralf Schützhold at the Technical University of Dresden in Germany and his colleagues have calculated that quantum fluctuations in the vibrations of a column of trapped magnesium ions are mathematically similar to the period of rapid expansion that followed the Big Bang.

Stretching and squeezing the column should cause the quantum fluctuations to become real vibrations, in much the same way that the stretching of space-time created matter and energy in the young cosmos. The group is now trying an experiment to see how well the analogy holds up.

JOURNAL CLUB

Arturo Zychlinsky
Max Planck Institute for
Infection Biology, Berlin,
Germany.

**A microbiologist wonders how
antimicrobial peptides beat
infection.**

My group is interested in why, although people often pick up infections and sometimes become ill, they almost always recover. Recovery is the result of a fantastically efficient immune system that relies in part on proteins

and peptides that kill microbes.

Antimicrobial peptides were discovered in systematic searches for potential drugs, and there are several types. Most are cationic and bind to the anionic surface of microbes. Recently, Roberto Lande at the University of Texas in Houston and his colleagues convincingly showed that one antimicrobial peptide, LL-37, can also bind DNA and serve as an activator for other immune cells (R. Lande *et al. Nature* **449**, 564–569; 2007).

LL-37 is one of several antimicrobial peptides that do

more than kill microbes and activate immune cells. Its other functions include chemoattraction and wound repair. But, as its name indicates, LL-37 has only 37 amino acids. It is plausible that its multiple effects on the host are due to its interaction with specific receptors that, in combination with other signals, result in diverse biological functions. More intriguing, however, is the fact that LL-37 kills microbes in the first place.

The question is whether LL-37 and other antimicrobial peptides truly function as bacterial killing agents in the host. Their

antimicrobial activity has so far been demonstrated only *in vitro*, where it might be a reflection of their cationic character. Another more attractive possibility is that infections are such an important threat to the host that, during evolution, many cationic proteins with diverse function were co-opted to serve as antimicrobial agents. If we do have many ways to kill microbes, maybe it is not so surprising that we often recover from infections.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS

Race to mimic human embryonic stem cells

Two much-anticipated scientific firsts announced this week bring the dream of regenerative medicine a step closer. The production of cloned primate embryonic stem cells and the reprogramming of adult human cells both represent important milestones in the quest to produce 'pluripotent' cells, which can develop into almost any of the body's roughly 200 cell types.

Human embryonic stem cells have this property, and those used in research are usually extracted from leftover embryos created during *in vitro* fertilization. But researchers want to create pluripotent cells that are genetically matched to individual patients. Such cells could then be transplanted to treat disorders such as Parkinson's disease and diabetes, or be used by researchers to model disease progression.

Cloning offers one way to create these cells. This week, a team led by Shoukhrat Mitalipov of Oregon Health & Science University in Beaverton report the first creation of embryonic stem cells from cloned monkey embryos (see page 497). Until now, cloned embryonic stem cells had been created only in mice. The accomplishment in primates is "like breaking the sound barrier," says Robert Lanza, of Los Angeles-based biotech company Advanced Cell Technology.

The creation in 1996 of the first cloned animal, Dolly the sheep, led to a rapid succession of clones. But continued failure to achieve cloned human or monkey embryos resulted in pessimism. In 2003, primate-cloning researcher Gerald Schatten said: "With current approaches, NT [nuclear transfer, a cloning technique] to produce embryonic stem cells in nonhuman primates may prove difficult — and reproductive cloning unachievable¹." This was after his study involving 716 monkey eggs failed to produce a single clone. Then, in February 2004, Woo Suk Hwang, then of Seoul National University in Korea,

announced that he had created cloned human embryonic stem cells². But, in January 2006, those results were shown to have been faked, and some suspected that Schatten was right.

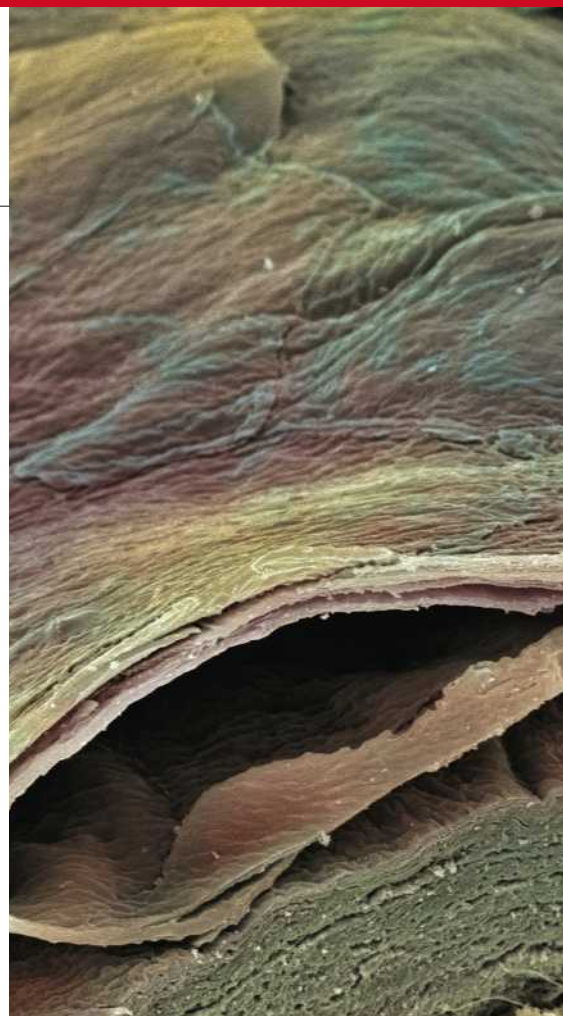
Mitalipov's group had been trying for almost a decade to achieve reproductive cloning in monkeys, and used some 15,000 eggs in the process. After Hwang's results turned out to be fraudulent, the researchers decided to switch from reproductive cloning to trying to establish a cloned embryonic stem-cell line. They took skin cells from Semos, a nine-year-old rhesus macaque, and inserted the cells' nuclei into eggs that had had their own genetic material removed. By January 2007, they had a cell line that retained its embryonic pluripotency — and, a couple of months later, another.

Mitalipov credits their success to a US\$19,000 imaging machine, named Oosight, which allows the structures in the egg that hold the DNA to be clearly seen, enabling easy extraction — the first step in nuclear transfer. Previously, researchers used a dye named Hoechst combined with ultraviolet light to locate and remove an egg's DNA. But Mitalipov's group found that this method damaged the egg.

Mitalipov says that his group's technique should work in human cells: "There's nothing specific here. But you must use this kind of imaging system," he says.

Reproductive cloning in monkeys, though, still looks to be a long way off. In April, after creating the two cell lines, Mitalipov's team tried transferring 77 embryos created into about a dozen surrogates. All failed to create pregnancies.

In the aftermath of the Hwang fraud, *Nature* has taken the unusual step of having these results independently verified by a team at Monash University in Melbourne, Australia³. "The scientific community has a need to have some certainty in the outcomes of nuclear-transfer experiments given the recent unfortunate experience with human somatic-cell nuclear



transfer. We have total confidence in these conclusions," says Alan Trounson, a member of the Monash team.

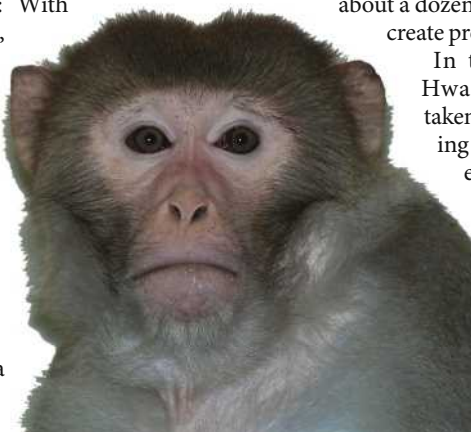
Many scientists are hesitant to apply the technique in humans because it requires women to undergo an uncomfortable procedure that involves significant health risks. Altogether, Mitalipov's team used 304 eggs to produce the two primate embryonic stem-cell lines. The researchers still have little idea of what separates the majority failures from the rare successes, so a similar number of eggs would probably be required to establish human lines.

Avoiding controversy

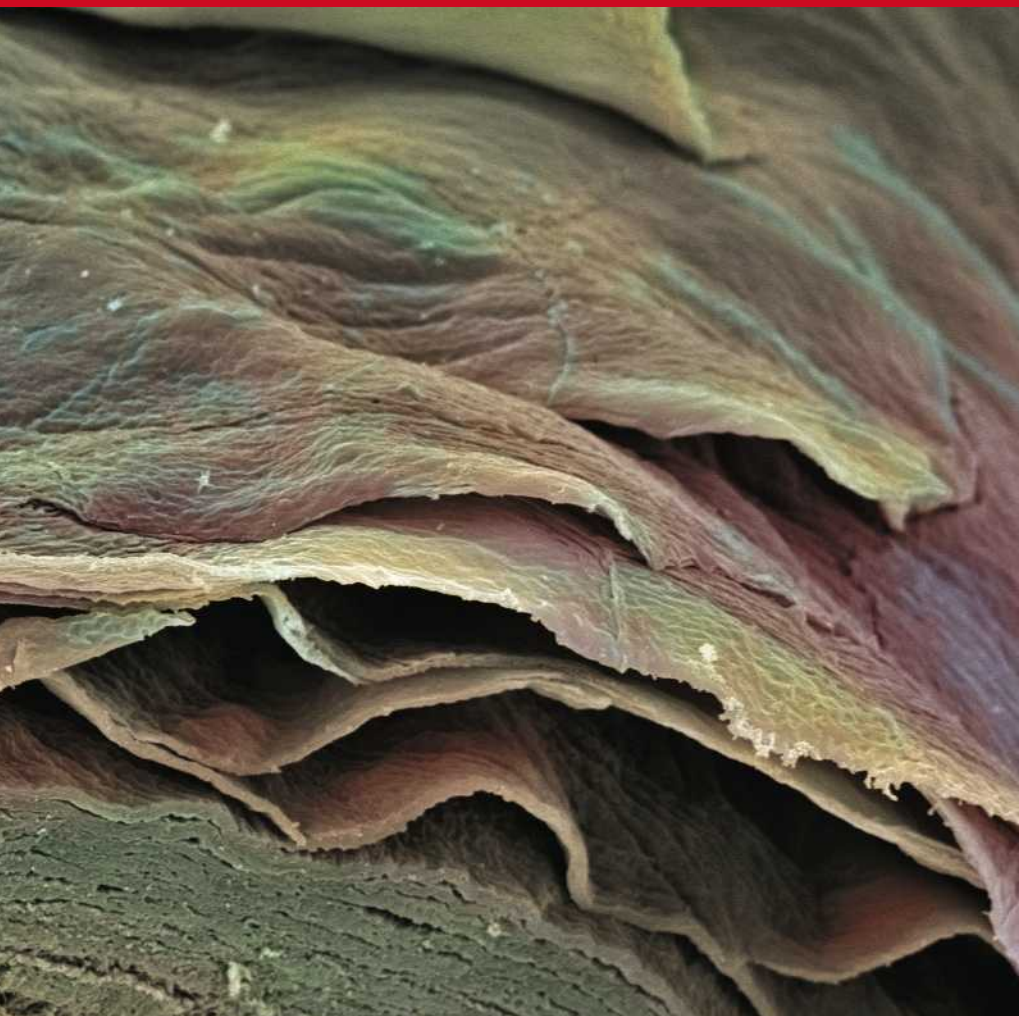
"The basic science of this is important," says John Gearhart, director of the Stem Cell Program at Johns Hopkins Medicine in Baltimore, Maryland. "But the low frequency of success here is troubling — particularly when considering human work."

But there is another promising route to creating pluripotent cells that does not require eggs or the controversial destruction of embryos. On Tuesday, Shinya Yamanaka of the University of Kyoto in Japan reported that his team had created pluripotent cells from human skin cells⁴ and, on the same day, a team of researchers led by James Thomson at the University of Wisconsin, Madison, reported the same⁵.

Yamanaka's work builds on his exciting



The donor cells used to create cloned primate embryos came from a monkey, Semos, named after the god in the science fiction work *Planet of the Apes*.



After reprogramming, cells taken from human skin (above) became embryonic-like stem cells.

discovery last year that introducing four transcription factors into mouse skin cells 'reprogrammed' the cells into an embryo-like state. Early this summer, Yamanaka and two other groups reported using the same four factors to create cells that seemed to be indistinguishable from embryonic stem cells⁶.

Because of the basic differences between human and mouse cells, Yamanaka was surprised to find that these four factors produced the same result in humans. The team generated 10 pluripotent lines from a culture of some 50,000 facial skin cells that had been subjected to the four factors. The skin sample, provided by a company in the United States, had been taken from a 36-year-old Caucasian woman. Yamanaka repeated the exercise with cells from synovial (joint) fluid from a 69-year-old man with similar results. During culture, the pluripotent cells take on the flat shape of embryonic stem cells.

Why so few cells successfully form such 'induced' pluripotent stem (iPS) cells is a mystery that Yamanaka is still trying to resolve. But because he uses such a cheap resource — cells that can be attained in their millions from a single skin biopsy — the low yield is not a problem, he says. In comparison to human embryonic stem cells — after much political wrangling and laborious effort, there are only

three cell lines in Japan — his technique is prolific, Yamanaka says. "In a little dish you can get ten cell lines quickly. Practically speaking it's a very high success rate."

Like the mouse iPS cells, Yamanaka's human iPS cells passed all the basic tests for embryonic stem cells, including the ability to form tumours expressing the three primary germ layers when injected under the skin of a mouse engineered to have no immune system.

But are they truly pluripotent? The most stringent tests, carried out in mice, are to see whether a whole individual can be created from iPS cells or whether iPS cells mixed with an embryo's are expressed in all of the resulting chimaeric mouse's tissues. Neither test can be done with human cells. "In humans, there's no answer to the question," says Yamanaka.

But, he adds, if the cells are for use in therapy or research on a disease affecting a particular tissue, it doesn't matter. Yamanaka's cells, for example, were able to form neurons, and cardiac muscle cells that — after differentiating for 12 days — started beating. But iPS cells do have drawbacks. Introduction of the four 'Yamanaka factors' requires genetic manipulation using viral vectors that health agencies would be unlikely to approve for clinical use.

And one of the factors, c-myc, is thought to be responsible for tumours in mice.

Thomson, who was the first to isolate and maintain human embryonic stem cells in culture, has gone part way towards solving these problems. He also used four factors, introduced by viral vectors, to reprogramme human foreskin cells. But only two of the four are the same, and he does not use c-myc. What is more, the discovery that a different recipe resulted in successful reprogramming suggests that scientists might have a greater degree of flexibility in finding clinically acceptable variations on Yamanaka's selection.

Onto the home strait

With researchers crowding into the field, Gearhart and others anticipate rapid advancement. "The iPS strategy is a major paradigm shift in reprogramming cells, and if proved effective and safe with human cells — most likely coming very very soon — it will diminish the role of somatic-cell nuclear transfer (SCNT) for deriving patient-specific pluripotent cells," says Gearhart.

The case for iPS-cell research and against using cloning was highlighted this weekend when the University of Edinburgh's Ian Wilmut, one of Dolly's creators, announced that he planned to turn his back on the field he pioneered in favour of research using Yamanaka's reprogramming technique.

Mitalipov maintains that the egg is the only "perfect reprogramming machine", and is confident that cloned cells will be the first to show therapeutic value. He says he has already more than doubled the efficiency of his cloning technique. And he is preparing for the clinic. Plans are underway to create cloned embryonic cell lines from Semos and some ten other monkeys,

to induce diseases such as diabetes in them, and then see if the cloned embryonic stem cells can be used to treat them.

Furthermore, in a move likely to raise fresh controversy, Mitalipov this week began a collabora-

tion with Alison Murdoch and Mary Herbert of Newcastle University, UK. Murdoch's group has a licence to work with human embryos. "I can't just keep modelling," Mitalipov says. "If we show medical progress, society will accept the technology."

David Cyranoski

1. Simerly, C. *et al. Science* **300**, 297 (2003).
2. Hwang, W. S. *et al. Science* **303**, 1669-1674 (2003).
3. Cram, D. S. *et al. Nature* doi:10.1038/nature06357 (2007).
4. Takahashi, K. *et al. Cell* doi:10.1016/j.cell.2007.11.019 (2007).
5. Yu, J. *et al. Science* doi:10.1126/science.1151526 (2007).
6. Cyranoski, D. *Nature* **447**, 618-619 (2007)

See Editorial, page 457.

"Until now, cloned embryonic stem cells had been created only in mice."

Drug firms accused of biasing doctors' training

Can the pharmaceutical industry be trusted to fund doctors' compulsory education without introducing bias? The issue is dividing Congress, academics and drugs companies. Now, preliminary data have emerged suggesting that industry-sponsored courses skew training material in favour of commercial interests.

Pharmaceutical firms spend over US\$1 billion a year to fund more than half of the continuing medical education (CME) courses that qualified physicians are required to take in the United States. Although drug firms say that they are scrupulous about separating their CME involvement from promotional activities, some may be influencing doctors, intentionally or not, and may even be putting patients' health at risk (see 'Smokescreens'). Drug companies deny this, insisting that the independent educational firms they pay to produce CME activities always operate outside industry influence. Government standards specify that commercial promotion must be kept separate from education.

Now, two small studies have attempted the first objective measurements of bias in doctors' education. Neither is large enough to settle the issue (and neither has yet been published), but their data suggest that industry sponsorship is distorting medical education.

Psychiatrist Jatinder Takhar, head of the CME office at the University of Western Ontario in London, Canada, first became interested in industry bias after attending a CME presentation on antipsychotic drugs that she had audited and approved previously. She was surprised to find that it did not match her recollection. "The data were slanted and the presentation was more promotional and less educational," she alleges.

Takhar and her colleagues went on to develop a standardized checklist of potential problems to be used for measuring bias in CME, which they published in June (J. Takhar *et al.* *J. Cont. Educ. Health Prof.* 27, 118–123; 2007). The team then applied its checklist to 17 company-sponsored CME events. Nine of these were found to be biased and should not have been approved, Takhar says. Some focused only on the sponsors' product and ignored rival treatments. In others, information on side effects associated with the sponsors' drugs was reduced to small print.

Another study, by Daniel Carlat, a psychiatrist at Tufts University School of Medicine in Somerville, Massachusetts, looked at printed

CME material — typically, medical articles followed by a written test. Carlat asked his colleagues to remove information about the sponsor from exercises sent to his office during 2005 and 2006. He then calculated the ratio of positive to negative statements made about every drug mentioned in the exercises. In 14 of the 15 exercises he looked at, the drug that received the highest ratio turned out to have been made by the firm sponsoring that exercise. He is preparing his paper for submission to the *American Journal of Psychiatry*.

Carlat publishes a CME newsletter that operates independently of industry funding, presenting a conflict of interest to his study. Researchers shown the findings of the two studies by *Nature* add that the samples Carlat looked at might not be representative of all written CME exercises. They also note that, unlike Takhar, Carlat has not validated his method for estimating bias. Drug companies also point out that firms tend to sponsor material about disorders for which they sell the best treatment, so it is right that those exercises focus on the sponsors' drugs. Yet interviews with physicians involved in the exercises that Carlat evaluated reveal that, at least in some cases, the focus is not simply the result of sponsors' drugs outperforming those of their rivals.

One exercise Carlat looked at was based on talks on bipolar disorder given at the 2005 US Psychiatric and Mental Health Congress. Anton



Porsteinsson of the University of Rochester Medical Center in New York gave one of the presentations. Porsteinsson alleges that he mentioned problems with a sponsor's drug — Depakote (divalproex), manufactured by Abbott Laboratories in Illinois — in his talk, but that these were missing from the exercise. Yet he claims that positive results about the drug that were not mentioned during the talk were

JOSE LUIS PELAEZ INC./CORBIS

Smokescreens

Some continuing medical education (CME) programmes have been accused of putting patient care at risk. One controversy centres on whether women who are pregnant should be tested for the herpes simplex virus. About one-quarter of women carry the virus, and it can harm their babies. But the chances of a mother infecting her child are very low. The extent to which treatment can protect babies is uncertain, and herpes drugs can cause side effects such as hypertension.

Organizations such as the US Preventive Services Task Force, part of the Department of Health and Human Services, therefore advise against screening for herpes in women who are pregnant.

That advice has not stopped drug giant GlaxoSmithKline from funding CME activities in which doctors advocate testing and, when necessary, treating pregnant women for herpes. If screening were introduced, the firm would potentially see a jump in sales of its herpes drug Valtrex (valacyclovir).

The CME programme "flies in the face of what expert panels recommend", says Adam Urato, an obstetrician at the New England Medical Center in Boston, Massachusetts. "It's about selling Valtrex." GlaxoSmithKline denies this, saying that it recognizes the expert advice and does not promote the use of its drug in pregnant women. The firm funds events, says spokeswoman Mary Rhyne, but the organizers select the speakers and set the content.

J.G.



Researchers are questioning the objectivity of some compulsory ongoing medical training for doctors.

included. In his opinion, “the big issue is that the article is selective”. Abbott denies claims that the presentation was distorted. A spokeswoman says that the firm had reviewed the material and found it to be in line with current thinking on bipolar disorder.

When asked about Carlat and Takhar’s findings, Jennifer Page, communications director at the Pharmaceutical Research and Manufacturers of America in Washington DC, said only that drug companies do adhere to her organization’s CME guidelines. Those standards require a firewall between pharmaceutical firms and medical-education companies. Drug companies can recommend speakers for events, for example, but CME providers select the physicians best suited for the job and are not obliged to follow a sponsor’s suggestions.

But the firewall is not working, some argue. As other forms of promotion are scaled back, marketing is increasingly disguised as education, says Jerome Kassirer, a former editor of the *New England Journal of Medicine*, now at Tufts University in Medford, Massachusetts. “It’s a reallocation of marketing money,” he says.

“CME providers know that if they don’t provide what the company wants then they’re never going to be hired again.”

One physician who was employed by a drug company to check the accuracy of promotional material, and who asked to remain anonymous, claims that CME material is still viewed by drug companies as part of marketing campaigns. He

adds that the provider–sponsor firewall is easy to subvert, because CME providers do not need to be explicitly told to produce biased programmes. “They know that if they don’t provide what the company wants then they’re never going to be hired again,” he claims.

The Accreditation Council for Continuing Medical Education (ACCME), based in Chicago, Illinois, which approves CME providers, acknowledges that more could be done to protect against bias. Murray Kopelow, the council’s chief executive, says that next year the ACCME will launch a trial in which reviewers will check CME material and report back on problems they encounter. The council will also consider whether drug companies could be asked to pool CME funds for distribution by an independent body in a bid to reduce the influence that any one firm has over an educational exercise. ■

Jim Giles

**HAVE YOUR SAY**

Comment on any of our news stories, online.

www.nature.com/news

UK 'terrorist' fights science-course ban

A British resident who is under surveillance for suspected terrorist activities is being prohibited from taking secondary-school-level science courses by the government, *Nature* has learned.

The man, referred to as A.E., is contesting the decision in court, in what is believed to be the first case of its kind. The preliminary hearing over whether A.E. should be allowed to take AS-level courses in human biology and chemistry took place on 16 November at London's High Court. The UK Home Office, which has an order restricting A.E.'s actions and affiliations, argues that such coursework could be turned towards terrorism. His solicitors counter that the knowledge is public, and that the furthering of A.E.'s education poses no threat.

At the heart of the case is a simple question: should basic courses in science be treated as potential tools for terror when in the wrong hands?

To protect the suspect, A.E.'s name and much of his personal information have been withheld from the public. What is known is that he is an unemployed Iraqi national in his mid-thirties who studied medicine at university in his home country. The government suspects him of terrorist affiliations, and he is the subject of a 'control order' — a special legal instrument that places limits on his freedoms.

Control orders were introduced by the UK government under the 2005 Prevention of Terrorism Act as a way of restricting the activities of suspected terrorists when prosecution would mean "revealing sensitive and dangerous intelligence". Among other things, control orders can be used to impose curfews, travel bans and limits on a person's access to mobile phones and the Internet. Nationally, 14 individuals are currently subject to control orders.

Critics have lambasted the orders as a blatant infringement of civil liberties. Because the orders depend on classified information,



A decision by the UK government to bar a suspected terrorist from taking a school science course is being contested at the High Court in London.

"There's never going to be an easy answer over issues of science and security."

subjects rarely know what evidence is being brought against them. Furthermore, the subjects are never charged with a crime. That makes control orders virtually impossible to rebut, according to Gareth Crossman, director for policy at Liberty, a London-based civil-rights organization. "The whole system is totally unfair," he says.

Since 2006, A.E. has been under a control order that has limited his movements and affiliations, according to his lawyer, Mohammed Ayub of Chambers Solicitors in Bradford, UK. The order has made it impossible for A.E. to find work, says Ayub. So he instead sought to further his education. English-language courses

went unopposed, but when in September A.E. applied to take the two science courses, the government told him he could not enrol.

A Home Office spokeswoman was unable to comment because of the ongoing court case, but Ayub says the government fears that A.E. could turn his coursework towards terrorist ends.

Scientists contacted by *Nature* say that AS-level coursework would be unlikely to help a terrorist. The human-biology material includes basic theories of disease transmission, according to Neil Roscoe, head of education and training at the Institute of Biology in London. But there's no information on how to intentionally spread pathogens. "I'd say that there's very little in the chemistry course that would help a terrorist act," adds Colin Osborne, who heads education at the Royal Society of Chemistry.

Still, both Osborne and Roscoe say the courses would include some information that might be useful to someone interested in committing a terrorist act. The human-biology course includes a detailed examination of neurotoxins, for example. And chemistry labs would contain some dangerous compounds and teach basic techniques such as distillation. "If the government wants to be cautious, there are aspects that could be considered as aiding the cause," says Roscoe. On issues

of science and security, he adds, "there's never going to be an easy answer".

But other scientists remain unconvinced. Peter Atkins, emeritus professor of physical chemistry at the University of Oxford, says the techniques taught in the course are simple: "Anybody with an interest in cooking could do them."

Ayub maintains that A.E.'s only interest in the classes is to return to university and continue his medical education. Ayub, himself a former oncologist, points out that A.E.'s level of knowledge is already well beyond an AS-level course. "It's absurd," he says.

The case will continue in the new year, with a ruling expected in the spring.

Geoff Brumfiel

RUPERT HORROX/CORBIS

WORDWATCH

Melody roads

At last, bored Japanese drivers have a way to amuse themselves. Engineers from Hokkaido have created roads with grooves at varying intervals, which produce a series of 'notes' as cars pass over them. It could even help cut speeding — the tunes apparently sound best at 45 kilometres per hour.

ZOO NEWS

Tiger trauma

Animal activists have slammed the Bing Chuan wild-animal park in Shenyang, China, after four of the park's tigers killed another and ate it. The attack has been blamed on inadequate food supplies for the animals, which usually get 5 kilograms of chicken or beef a day. Cold weather has caused revenue from visitors to dwindle, say officials, and rations for the animals have dwindled with them.

3 GOOD REASONS...

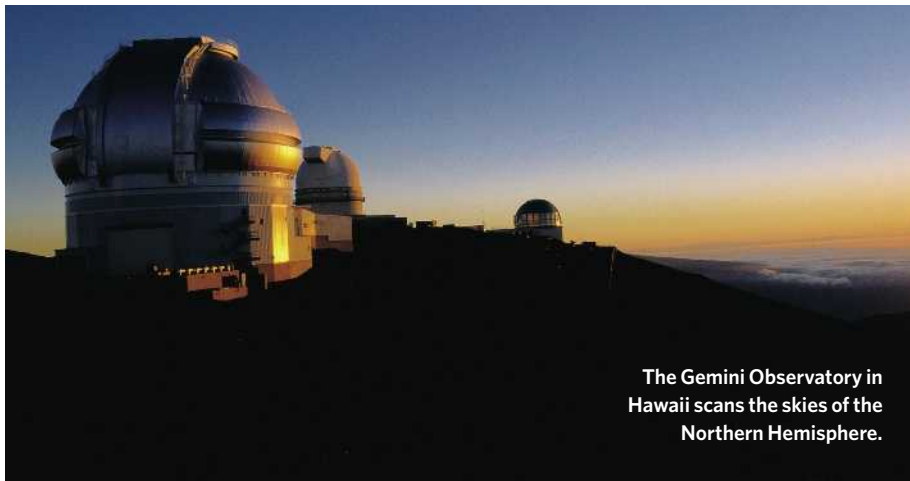
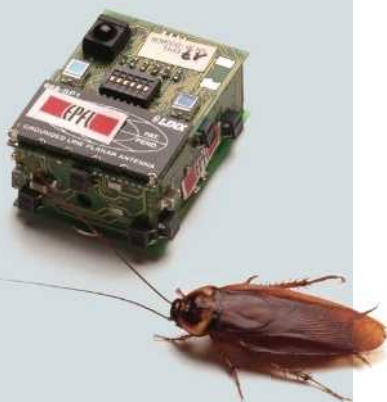
To let the robots take over

1 An end to the misery of cockroach infestation: Swiss and Belgian roboteers have created a 'Pied Piper' robot (pictured below) that can influence the insects and 'herd' them into a desired area.

2 Fun theme parks: South Korea is to build two robot theme parks by 2013, allowing people to meet and interact with their electronic counterparts.

3 Freedom from wrist pain: a German calligraphy-bot has been programmed to write out the entire Martin Luther Bible.

Sources: Guardian, AP, Science, Networkworld.com, Sawse.com



The Gemini Observatory in Hawaii scans the skies of the Northern Hemisphere.

UK astronomers stunned by Gemini withdrawal

The United Kingdom has abruptly announced its intention to withdraw from the Gemini Observatory, potentially leaving hundreds of British astronomers without a major telescope in the Northern Hemisphere.

The Science and Technology Facilities Council (STFC) said on 15 November that it would pull out of the international observatory, with its twin 8-metre telescopes in Chile and Hawaii. The STFC, one of seven UK national research councils, pays for major physics and astronomy projects. It has nearly a quarter stake in Gemini and pays annual subscription fees of £4 million (US\$8.2 million).

The British astronomical community has reacted with shock and dismay. "It's a mistake," says Royal Astronomical Society president Michael Rowan-Robinson. "I think it damages our ability to do multi-wavelength astronomy."

UK astronomers were not consulted before the decision, says Roger Davies, an astrophysicist at the University of Oxford and former Gemini board chair. "I was shocked," he adds. "Gemini is a world-ranked facility — it's our only access to the Northern Hemisphere."

Astronomers in the United Kingdom can still use the Southern Hemisphere telescopes of the European Southern Observatory in Chile. Rowan-Robinson wants the STFC to negotiate a partial partnership with Gemini that would allow them to retain observation time in Hawaii.

The Gemini observatories opened in 1999 (north) and 2001 (south), with capabilities

in both the optical and infrared wavelength ranges. The infrared allows astronomers to peer deep into the ancient corners of the Universe. As part of the Gemini Deep Deep Survey, astronomers discovered that young galaxies, forming 3 billion to 6 billion years after the Big Bang, were bigger and more mature than previously thought (K. Glazebrook *et al.* *Nature* **430**, 181–184; 2004). Other Gemini member countries include the United States, Canada, Australia, Argentina, Brazil and Chile.

The STFC announced its intention at the Gemini board meeting in La Serena, Chile. On

21 November, it must inform its own governing council on how it plans to save £80 million in the next three years. "In this financial climate, we are staring very hard at our programmes," says spokesman Peter Barratt.

The STFC paid £30 million between 1994 and 2001 towards Gemini's construction. By leaving the observatory, the council would pay a penalty of £8 million. UK astronomers would immediately forfeit their observation time, which was 23% of the available time on each telescope, or about 70 nights of the year.

The Gemini board issued a statement on 16 November saying it was confident that it could resolve the funding without interrupting operations. Wayne Van Citters, a Gemini board member from the National Science Foundation, the US funding body, noted that the STFC had not yet formally given notice. There could still be room to negotiate, he says.

Eric Hand

"In this financial climate, we are staring very hard at our programmes."

UNIV. OXFORD

J. HALLOY



ENDANGERED PORPOISE WORSE OFF THAN THOUGHT

Mexican marine mammal
may have two years left.

www.nature.com/news

C. FASEI/PROYECTO VAQUITA/O. VIDAL/WWF-MEXICO

Proposal raises bones of contention

Alarm is growing among anthropologists in the United States over a plan that could empty institutions of about 120,000 human skeletons currently stored for research purposes.

Under a new proposal, the bones at museums, universities and federal facilities across the nation could be given to Native American tribes now living in the area from which the remains were excavated, even if the skeletons are not culturally identifiable to the tribes. In October, the Native American Graves Protection and Repatriation Act (NAGPRA) programme, the agency that oversees the handling of American Indian remains, opened a 90-day comment period on the proposal.

It would affect ancient skeletons similar to the Kennewick Man specimen from Washington state. Scientists won a long court battle to keep that 9,000-year-old skeleton for study after attempts to give it to tribes for likely disposal.

Under current guidelines, bones that can be matched with a living tribe by using cultural

evidence may be returned to them, but what to do with remains that are so old that they can't be associated with today's surviving tribes has been hotly debated for a decade.

Setting up a system to dispose of these 'culturally unidentifiable' specimens is a natural progression from NAGPRA, which became federal law in 1990 after extensive negotiations between scientists and tribes, says Sherry Hutt who manages the NAGPRA programme. "This is a proposal for disposition [in situations] such as when a new highway runs through an old graveyard," says Hutt, an attorney and economist from Arizona.

But major scientific organizations strongly dispute this view, calling the move "illegal" because it goes beyond the Congressional law, and a "divisive" manoeuvre that may shatter decades of working relationships between scientists and tribes. "The rules

would be disastrous," says Phillip Walker, an anthropologist at the University of California, Santa Barbara. A former member of NAGPRA's seven-person review committee, Walker helped prepare the American Association of Physical Anthropologists (AAPA) comments.

The AAPA says that the proposal would allow the return of almost any skeleton, even those used in medical schools, would greatly hinder anatomical teaching, and would eliminate comparative material for studies. Many of the specimens are among the oldest, offering data on the continent's first humans.

The comment period closes on 14 January, when the NAGPRA review committee will comment on the draft regulations. Then US Secretary of the Interior Dirk Kempthorne will issue a final decision.

Rex Dalton

**"The proposal would
allow the return of
almost any skeleton,
even those used in
medical schools."**

Germany sets up national academy of sciences

The Leopoldina, a 350-year-old scientific academy based in Halle, is to become Germany's first national academy of sciences.

The country's science minister, Annette Schavan, announced her decision in a radio interview on 16 November — without having first told the Leopoldina or Germany's seven regional science academies. In its new role, the Leopoldina will advise the government, parliament and the public about socially relevant scientific issues.

Schavan hopes that the decision will end a decade-long debate over whether to promote any of the regional academies to the national level. Germany's scientific organizations, including the Leopoldina, had recommended strongly against the move — favouring instead the creation of a 200-strong council appointed by the regional academies.

The 1,280-member Leopoldina now plans to collaborate with the other academies to produce science-led policy advice, says its secretary-general, Jutta Schnitzer-Ungefug. See Editorial, page 458.

Climate body's summary urges action on warming

The final summary of this year's three massive reports from the Intergovernmental Panel on Climate Change (IPCC) contains notably more urgent language than before.

Released on 17 November in Valencia, Spain, the synthesis report talks about "dangerous anthropogenic interference with the climate system". That phrase was absent from previous IPCC publications because of objections from some political delegates to the panel's conferences.

The report confirms that it is at least 90% certain that global warming is real



IPCC chairman Rajendra Pachauri (left) and UN secretary-general Ban Ki-moon at the release of the climate body's synthesis report.

and caused by human activities, and sets out possible strategies for avoiding global temperature rises that could reach as high as 6°C by the end of the century.

Hans Verolme, head of the WWF Global Climate Change programme in Washington DC, points to a section called 'reasons for concern', which he says acknowledges the risk of both abrupt climatic shifts and irreversible impacts. Although there is "a little hedging and qualifying going on", he says, the fact that such a statement is in the report is notable.

Stem-cell researcher accused of negligence

An investigation by an Australian university has found evidence of negligence but not fraud in a project headed by stem-cell researcher Alan Trounson. The focus has been on an unnamed senior researcher. Trounson, incoming president of California's US\$3-billion stem-cell research programme, has not been accused of misconduct.

Monash University in Victoria launched the investigation in April, after an audit by the Australian Stem Cell Centre found inconsistencies in quarterly progress reports submitted by the research team. The Aus\$1-million (US\$890,000) project, which aimed to find stem-cell therapies for smoking-

induced lung damage, has been on hold since February, when questions first arose.

Investigators concluded that the senior researcher, who has left the university, lacked sufficient management experience for his position, and was negligent in his reporting of the data. Monash is considering whether to repay the funds to the stem-cell centre.

Presidential veto leaves NIH facing shortfall

Funding for the US National Institutes of Health (NIH) is in limbo as President Bush and Congress fight over the country's financial priorities. On 15 November, the House of Representatives failed to override a presidential veto of a spending bill that includes the NIH along with other health, education and labour programmes.

The action means that the \$28.9-billion budget of the NIH will not get a proposed 3.1% increase in 2008, putting the agency on the road to a far smaller increase or even a cut from its 2007 funding level. The House fell two votes short of the two-thirds majority needed to override a veto.

House appropriations chairman David Obey (Democrat, Wisconsin) says that shaving billions from the huge bill to avoid another presidential veto will mean cutting at least \$700 million from what would have been a \$1.1 billion increase for the NIH.

Midwest coalition joins fight against emissions

Six US governors and the premier of Manitoba in Canada have agreed to cut greenhouse-gas emissions and set up a midwestern carbon-trading market, marking the third such regional accord.

With similar pacts in place in the northeastern and western states, roughly half the US population is now covered by some form of greenhouse-gas agreement, according to the World Resources Institute. The governors of Wisconsin, Minnesota, Illinois, Iowa, Michigan and Kansas signed the accord on 15 November. South Dakota, Indiana and Ohio will be observers, allowing them to help set up the carbon trading.

Meanwhile, a US appeals court has thrown out fuel-efficiency regulations instituted for sports utility vehicles (SUVs) and other light trucks by the Bush administration last year. In its 15 November ruling, the court in San Francisco ordered the administration to strengthen the standards and address CO₂ emissions.

Correction

In the Graphic Detail in last week's News (Nature 450, 327; 2007), the label South Korea erroneously pointed to mainland China.

Congolese government creates bonobo reserve

The Democratic Republic of the Congo is to establish the world's largest protected area for great apes. Larger than the state of Massachusetts, the Sankuru reserve is targeted to save bonobos (*Pan paniscus*; pictured), one of the two species of chimpanzee.



An estimated 10,000 bonobos remain and are found only in the Congo. The apes are often hunted for bushmeat, and conservation efforts fell by the wayside during the Congo's latest civil war. The new, 30,500-square-kilometre reserve — established with the help of the Bonobo Conservation Initiative in Washington DC — brings the amount of the Democratic Republic of the Congo set aside for wildlife to just over 10%. The government's target is 15%.

BUSINESS

Trip into the unknown

A plant in Uganda hopes to sell cut-price drugs by taking advantage of exemptions from rules that protect patents. But its operators face major obstacles, as **Tatum Anderson** reports.

On the outskirts of Kampala, and rather at odds with the dusty, low-rise buildings around it, stands a state-of-the-art, glass-fronted pharmaceutical plant.

Opened with great fanfare last month by Yoweri Museveni, the president of Uganda, the factory will manufacture drugs to treat HIV/AIDS and malaria for markedly less than the cost of imported drugs. The World Health Organization (WHO) estimates that in Uganda, fewer than 38% of the patients with HIV/AIDS that need antiretroviral drugs receive them. But although Africa carries the largest burden of HIV/AIDS and malaria in the world, it produces only 1% of the world's drugs, according to the World Bank.

"We want local solutions to local problems and this is the first step to say we can do it ourselves," says Fred Kitaka, finance director of Quality Chemicals, the Ugandan company that co-runs the plant.

Previously a drug importer, Quality Chemicals put up two-thirds of the US\$30-million it cost to build facility, which can produce 6 million tablets per day. Cipla, a drug manufacturer based in Mumbai, India, has provided the remaining US\$8.5 million, as well as consulting services. The Indian firm also holds 50% of the shares in Quality Chemical Industries, the company that runs the new plant.

The factory will start by producing some of Cipla's most widely used fixed-dose combination drugs, including an antiretroviral called Triomune, which contains lamivudine, stavudine and nevirapine. It will also produce Lumar-tem, an antimalarial combination therapy that is made by Novartis under the brand name of Coartem. The plant will start manufacturing early next year and will initially sell antiretroviral drugs at two-thirds of the cost of imported products. As production volumes rise, the factory claims that it will reduce prices further.

Kitaka says that the company can produce cheaper drugs by taking advantage of the staggered timetable for compliance with Trade-Related Aspects of Intellectual Property Rights (TRIPS), a global agreement under which companies can enforce patents on their products and

prevent copycat producers from selling them.

At the moment, many of the drugs used by Ugandans are manufactured in India. But changes to India's patent law in 2005 to bring it in line with TRIPS mean that newer drugs are more likely to be patented, and therefore more expensive, says Kitaka. As one of the world's least-developed countries, Uganda does not need to pass TRIPS-compliant laws until 2016. It can therefore manufacture drugs that would

be patented elsewhere until that time. (As it happens, Uganda is planning to revise its patents laws earlier than that.)

Figures from the WHO indicate that by 2005, 37 African countries had pharmaceutical industries, albeit on a small scale. There is strong political will in Africa to boost these facilities. But a 2005 World Bank study has shown that local manufacturing often does not make economic sense for poor countries. Production costs rise considerably because African countries often lack reliable supplies of electricity and water, as well as the engineers, chemists and other skilled staff needed to run drug factories. Raw materials for the drugs, as well as plastics and other packaging materials, may also have to be shipped in from elsewhere.

Additional, plants in India and elsewhere that manufacture low-margin antiretrovirals often share overheads with lucrative, high-margin drugs. An African plant, however, "would have to recover all costs, including salaries, exclusively from the antiretrovirals, which may make it unviable," Juneja says.

The effect of TRIPS exemptions on countries such as Uganda are also limited by the fact that so many of the drugs imported from India are off-patent anyway, observers say. And even though Africa is home to some highly reputable manufacturers, such as South Africa's Aspen Pharmacare, which makes drugs licensed by GlaxoSmithKline, many health-care officials have doubts about the quality control of drugs produced in other African facilities.

Factories must reach international manufacturing standards before aid agencies will buy their drugs or other countries will import them. But the WHO, which certifies factories as having reached such standards, says that not many African manufacturers have even applied for inspections.

Industry observers say that with few inspectors to certify plants, factories wait perilously long for inspections. "We are finding high-quality local producers who are going out of business because the certification process is not resourced to do the job," says Philippa Saunders of Oxfam's Essential Drugs Project, based in London.

Despite the odds, Uganda's government, manufacturers and Cipla are determined to surmount these obstacles at the Kampala factory. The government — which helped to broker the deal between Cipla and Quality Chemicals — plans to expand Uganda's processing of active ingredients so that raw materials do not have to be imported. And in a technology-transfer arrangement, Cipla has overseen the design and building of the Quality Chemicals factory so that it meets international standards and can export to other African countries. It is also providing 40 staff to train Ugandan workers to run the factory over the next couple of years.

So Kitaka is confident of a return on investment in six or seven years, although he acknowledges that the firm's success will depend on a great many factors. And the stakes are more than financial. "We're not just making money," he says, "we are helping the communities we live in." ■



President Yoweri Museveni gives a thumbs up to Uganda's latest drug venture.

be patented elsewhere until that time. (As it happens, Uganda is planning to revise its patents laws earlier than that.)

Figures from the WHO indicate that by 2005, 37 African countries had pharmaceutical industries, albeit on a small scale. There is strong political will in Africa to boost these facilities. But a 2005 World Bank study has shown that local manufacturing often does not make economic sense for poor countries. Production costs rise considerably because African countries often lack reliable supplies of electricity and water, as well as the engineers, chemists and other skilled staff needed to run drug factories. Raw materials for the drugs, as well as plastics and other packaging materials, may also have to be shipped in from elsewhere.

Viable markets

Sandeep Juneja, HIV project head at the Indian manufacturer Ranbaxy Laboratories says he doubts that many plants will be viable in Africa. An export market that consists purely of other low-income countries would be too small, he says. And under TRIPS, compa-



LIGHTS IN THE DEEP

Far below the surface of the ocean, beyond the reach of the Sun's rays, organisms still have eyes. **Mark Schroppe** investigates seeing without sunlight.

It is a calm August day in the azure waters of the Bahamas, and Erika Raymond, a doctoral student in Oceanography at Johns Hopkins University in Baltimore, Maryland, has been screening hours of video footage inside a shipboard laboratory. The footage comes from a camera system that was stationed on the sea floor some 600 metres down.

Hardly a trace of sunlight makes it that deep, and trying to film in the darkness is a tricky feat. Artificial lights might scare creatures away, or attract the wrong kind. So the camera is fitted with a red light, thought to be invisible to the eyes of most deep-sea creatures. But the system does more than film the depths. Eye-in-the-Sea, as it is called, is equipped with an LED lure, a cluster of tiny lights designed to flash and flicker in very specific patterns, all in the hope that something might respond.

Raymond is getting excited about what is on the screen. She calls over Edith Widder, her PhD adviser, who becomes similarly animated. "That is so great," says Widder, co-founder of the Ocean Research and Conservation Association in Fort Pierce, Florida. Others gather round, expecting to see another good clip of one of the giant sixgill sharks that have been nosing around the rig. But the duo's

excitement stems from something more subtle and more profound.

First there's a repetitive flash of light from the lure in the centre of the screen. Then, in the distance, there's a similar burst, but this one is from an animal. Raymond and Widder jump forward to other clips in which the lure was making that same flashing pattern, one of five they have been using. Again they see the response, sometimes multiple responses. It is official. Through their lure, the scientists have begun to communicate with deep-sea organisms using what might be the loudest form of expression in that dark void, the language of bioluminescence — biochemically produced light (see graphic opposite).

What exactly had been said is another issue entirely. The response could have been the equivalent of a flirtatious wink or, more likely, a warning call. But simply making that connection for the first time is a good start in a field of study that, of necessity, operates in the dark.

True behavioural observations of bioluminescent activities in the deep are rare, because of the difficulties inherent in keeping deep-sea species alive in the lab, and the severely limited access to the depths of the oceans. Moreover, the tools of on-site exploration — remotely

operated vehicles and submersibles — generally fall short of being unobtrusive. Ron Douglas, who studies deep-sea vision at City University in London, compares exploration by submersible to taking a "Land Rover and going out into the savannah in the middle of the night with the stereo on full blast, the lights on full, with a rotating siren and expecting to see normal lion behaviour".

Plumbing the depths

More restrained research such as that conducted this summer in the Bahamas, the third in a series of expeditions funded by the National Oceanic and Atmospheric Administration's Office of Ocean Exploration, is gradually revealing new information about what light can be found in the depths, how the organisms there use it, and even how they see their world. Although much uncertainty remains, one thing is abundantly clear: bioluminescence is nearly everywhere, and for inhabitants of the deep sea, it seems to play key roles in everything from eating, through mating, to staying alive. Tantalizing recent discoveries also suggest that at least some deep-sea organisms are seeing more than anyone expected.

Descending into the crystal waters of the

E. WIDDER

Bahamas, observers are invariably struck by the remarkable blues not only of the water, but of everything in it. The long wavelength red light is quickly absorbed and extinguished by the uppermost layers of water. This is why most deep-sea creatures don't see red — there is hardly any of it. Travel a bit deeper, say 200 metres or so, and there is still enough light for a passable twilight. But as surface light fades, the first small flashes of bioluminescence from dinoflagellates appear. Go deeper still and the flecks become more prevalent. Larger flashes, perhaps from a shrimp or a jellyfish, punctuate the scene. Eventually, the lightshow grows into a veritable fireworks display against an ever blacker background. By 500 metres or so humans can't detect much if any sunlight. The animals equipped for this realm might still detect some light below that level, but at 1,000 metres, absolutely every trace of sunlight is extinguished.

Seeing without sunlight

Yet even beyond sunlight's reach, eyes are still common. Bottom-dwelling creatures that spend their entire lives shielded from the light of the Sun actually tend to have enlarged eyes. Those eyes have to be seeing something and that something is the bioluminescence. Indeed, ongoing surveys of bioluminescence led by Monty Priede, from the University of Aberdeen, UK, show that although its frequency decreases with depth, bioluminescence persists thousands of metres beneath the surface.

"People lose sight of the fact that light is the most important variable in our environment, and it is probably as important in the marine environment," says Widder, "but we have to understand they're not seeing like we do."

Researchers suggest that although other senses are clearly important, some deep-sea organisms depend on bioluminescence for every life function. Not surprisingly, given the scarcity of hiding places, one of the most common uses is simply survival.

One established theory is that certain organisms use bright bioluminescent flashes like burglar alarms to startle predators, or to attract their predators' predators¹. Widder says that one of the earliest deployments of Eye-in-the-Sea may support the second option. In 2004, just more than a minute after the lure was activated in a jellyfish-like pattern for the first time at depth, a 2-metre squid arrived on scene. The next year, hundreds of miles away, a squid of the same as-yet-unidentified species responded similarly. Widder says that it was

terrific proof that unobtrusive observation could capture novel behaviour. "I don't think anything can top that squid," she says.

Circumstantial evidence suggests that deep-sea animals have other uses for their bioluminescent organs, known as photophores, as well as bioluminescent tissues and various forms of 'spew'. Sometimes it's a deterrent. Some organisms seem able to light up a predator's stomach when they are eaten, broadcasting its location to other predators². Bioluminescence similarly could be used like the colours on a poisonous snake or frog to warn would-be predators of toxicity or unpalatability³. On the 2005 Deep Scope expedition, Widder discovered the first known bioluminescent anemone, which produces a sort of glowing slime. Its use is not known, but she speculates the slime could be just such a warning. There is even evidence of batesian mimicry, where perfectly edible organisms have evolved to impersonate noxious species.

Many animals can hide from predators using light. In the ocean's twilight zone, where some light still penetrates, a dark silhouette is easily spotted from below, and is hence a dangerous liability, says Sönke Johnsen, a biologist at Duke University in Durham, North Carolina, and the 2007 Deep Scope chief scientist. So, many animals use photophores to counterilluminate, or light up their undersides and blend⁴.

On the flip side of bioluminescence, there is strong evidence that many animals use light to attract and find others. Lots of fish and squid, for instance, use bioluminescence as a form of searchlight. Others attract mates using sex-specific bioluminescent patterns⁵. And female deep-sea anglerfish are famous for their lures, filled with bioluminescent bacteria, that they dangle in front of imposing fangs.

Bioluminescent bacteria are common throughout the ocean, and researchers have proposed that the bacteria use their luminosity to find good homes. Bacteria that are lit up, whether hitching a ride on a bit of detritus or on a fecal pellet, are more likely to be eaten by fish, in whose guts their needs are well met.

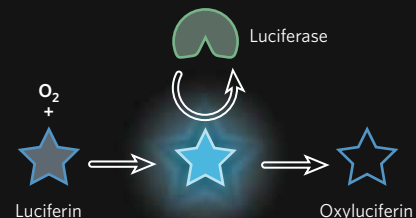
Look deep into the eyes

Understanding what deep-sea animals can see is, of course, integral to understanding how bioluminescence might be used and to what degree of success. Tammy Frank, a Deep Scope leader and a visual ecologist at the Harbor Branch Oceanographic Institution in Florida, which owns the ship and submersible used, has done extensive work studying deep-sea

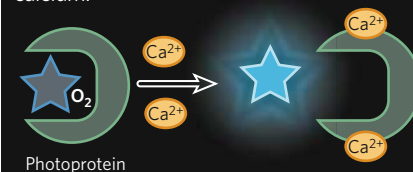
Living light: how it works

Bioluminescence — biochemically produced light — exists in many organisms, but the vast majority is in marine environments.

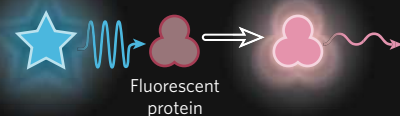
Different chemistries exist, but both on land and in the sea, the reaction takes place when a substrate, generally called a luciferin, is oxidized by a catalyst called a luciferase. Luciferin substrates can be created internally by the organism or possibly obtained through the diet.



Substrate, catalyst and co-factors such as oxygen can be bound together in a single unit — a photoprotein — and triggered to produce light by the introduction of an ion, typically calcium.



Fluorescence, however, is when light is absorbed and re-emitted rather than being chemically produced. When a fluorescent protein is excited by one type of light, such as that produced by bioluminescence, light is emitted at a longer, lower-energy wavelength.



Fluorescent proteins such as green fluorescent protein, isolated from the jellyfish *Aequorea victoria* are used extensively in biomedical research, as is bioluminescence. Calcium-activated photoproteins can be used to study muscle contraction and other signalling. And luciferin/luciferase combinations can be used to report the activity of drugs, genes and cells.

"Light is the most important variable in our environment, and it is probably as important in the marine environment."
— Edith Widder

eyes. She speculates that the size of the eyes of bottom-dwelling deep-sea creatures, relatively larger than those of the occupants of the water column above them, could be tied to energy expenditure. The advanced vision requires energy. Creatures in open water cannot afford to waste energy that might be better spent outswimming predators. Bottom-dwelling animals can use the sediment or rock structures to help them hide, which could allow them to funnel more energy towards vision.

To date, most deep-sea animals studied seem able to see light only in the blue-green range. These shorter wavelengths penetrate the water farther and are where the majority of bioluminescence falls. But there are notable exceptions. In 2005, on the second Deep Scope expedition in the Gulf of Mexico, Frank discovered a bottom-dwelling crab species sensitive not only to the standard blue-green range, but also to ultraviolet light. The utility of this ability is not clear, but could, by some unknown means, allow the crab to detect its favoured hiding spots — soft corals — or to distinguish between different types of bioluminescent light, which does at times dip very slightly into the ultraviolet range. If such an ability is proven, it would not be the first time scientists have discovered visual skills that seemed initially bizarre or improbable.

Because of red light's weak ability to travel in water, conventional wisdom had long held that all bioluminescence in the deep was blue or green. But in 1981 Peter Herring, a now semi-retired bioluminescence pioneer at the University of Southampton, UK, reported the strange discovery of red bioluminescent 'searchlights' under the eyes of certain dragonfish⁶, a common family of elongated deep-sea fish with menacing fangs appropriate to their name (see picture page 472). Herring, and others, have shown that these searchlights are the result of fluorescent proteins that shift bluish bioluminescence to the red range. The eyes of these fish are adapted to see red light as well, suggesting the animals can secretly attract mates or hunt using flashlights that few other fish can see.^{7,8}

Strangely, one dragonfish species initially did not seem to have the pigments



Edith Widder with the Eye-in-the-Sea submersible.

necessary to see the red light. Douglas and his colleagues eventually discovered that these fish use a chlorophyll pigment from bacteria for red vision. There are still significant uncertainties about how the pigment is obtained and how it accomplishes the energetic shift that leads to red vision. But other groups have confirmed that when the chlorophyll is added to the eyes of animals such as mice, it boosts red vision⁹.

Seeing red

Steven Haddock, at the Monterey Bay Aquarium Research Institute in Moss Landing, California, and others have proposed that there may be still more uses for red light in the deep. In 2005, his group made the controversial proposal that at least one fish-eating species of siphonophore, a jellyfish relative, combines bioluminescence with red fluorescent proteins to create red fishing lures on its tentacles¹⁰.

Challenges to the theory revolve around the low number of deep species so far known to see red, although the vast majority have not been tested. Physics provides another challenge. Fluorescence is inefficient, because energy is lost in the transfer from bioluminescent reactions to fluorescent proteins, meaning a relatively weak signal is produced or received.

"My physics head says, 'No,'" says Justin Marshall, another Deep Scope participant, from the University of Queensland in Brisbane, Australia, "But my biology head says, 'Well, Why not?' Biology is weird, so it could be."

Haddock, and Mikhail Matz, another Deep Scope leader, based at the University of Texas in Austin, are involved in a related research project focused on the red fluorescent protein found in

the dragonfish searchlight photophores, and, according to preliminary results, also in the chin barbels of two species. There it is not associated with a bioluminescent organ, making its presence all the more intriguing. "I think red fluorescence attracts some sort of prey," says Matz, but what and how are not clear. The dragonfish protein seems to be novel, and, as with several collected from shallower-dwelling animals, Matz is exploring the potential use of the protein as a new tool for biomedical research, where other fluorescent proteins are used frequently.

Questions such as the extent to which red fluorescence might play a role in the deep are likely to remain unanswered for some time, but Widder hopes that there will soon be a new tool to advance research on a number of the deep's mysteries. Thanks to military funding, most research on bioluminescence has focused on shallow waters where it can reveal the presence of submarines and other vessels. But Widder has now secured funding from the National Science Foundation for a new version of Eye-in-the-

"My physics head says 'No' but my biology head says, 'Well, Why not?'"

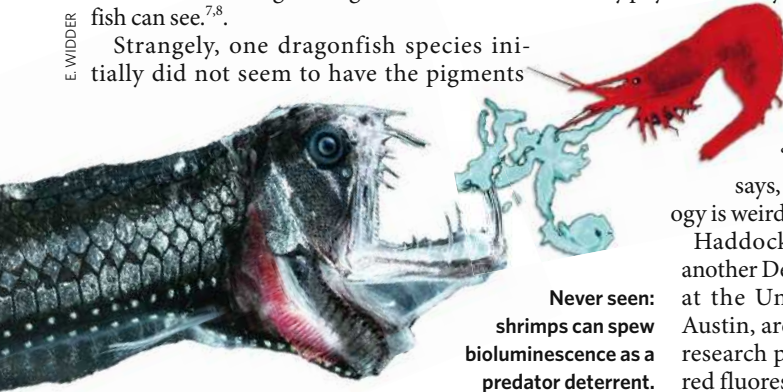
— Justin Marshall

Sea, scheduled for deployment as part of an undersea observatory array off Monterey, California, in early 2008. This system will provide researchers with the first ever

perpetual, unobtrusive view of the depths. That could mean the first effective, long-term study of true deep-sea bioluminescent behaviour. Among numerous other potentials for discovery, this extended view may allow Widder and her colleagues to decode just what was said between their lure and the unidentified creatures they were communicating with in the Bahamas. "I feel like we'll be able to address some of these issues for the first time," she says.

Mark Schroppe is a freelance writer in Florida.

1. Widder, E. A. in *Adaptive Mechanisms in the Ecology of Vision* (eds Archer, S. N., Djamgoz, M. B. A., Loew, E. R., Partridge, J. C. & Vallerga, S.) 555–581 (Kluwer Academic, Boston, 1999).
2. Johnsen, S. *Biol. Bull.* **201**, 301–318 (2001).
3. Grober, M. S. *Anim. Behav.* **36**, 493–501 (1988).
4. Johnsen, S., Widder, E. A. & Mobley, C. D. *Biol. Bull.* **207**, 1–16 (2004).
5. Herring, P. J., *J. Mar. Biol. Assoc. UK* **87**, 829–842 (2007).
6. Herring, P. J. in *Bioluminescence and Chemiluminescence*, (eds DeLuca, M. A. & McElroy, W. H.) 527–530 (Academic Press, New York, 1981).
7. Widder, E. A., Latz, M. F., Herring, P. J. & Case, J. F. *Science* **225**, 512–514 (1984).
8. Douglas, R. H. *et al. Nature* **393**, 423–424 (1998).
9. Washington, I. *et al. Photochem. Photobiol. Sci.* **6**, 775–779 (2007).
10. Haddock, S. H. D., Dunn, C. W., Pugh, P. R. & Schnitzler, C. E. *Science* **309**, 263 (2005).



**Never seen:
shrimps can spew
bioluminescence as a
predator deterrent.**



INNATE ABILITY

With keen immunological insight and a knockout mouse 'factory', Shizuo Akira leads by quiet example. **David Cyranoski** visits the world's most-cited scientist as he prepares to run one of Japan's premier research centres.

In the spring of 2000, Osamu Takeuchi and Hiroaki Hemmi were scanning row upon row of tiny plastic wells filled with a yellow fluid. Finally they found one that had remained clear. The fluid came from cultures in which mouse immune cells that lacked a specific receptor had been challenged with different components of infectious agents — bits of bacterial cell wall, DNA and the like. Yellow signalled that the cells had reacted to the components. Clear meant that they hadn't, revealing the purpose of the missing receptor. The clear well meant that the pair had discovered the function of toll-like receptor 9 (TLR-9), one of a family of proteins that have revolutionized the field of immunology. "I knew we had something right there that no one else in the world knew about," recalls Takeuchi. "I knew it would be a big article."

The article¹, published in 2000, would be cited in nearly 2,000 other articles, according to Thomson Scientific's Web of Science. Yet Takeuchi's excitement at their discovery was somewhat lost on his mentor, Osaka University's Shizuo Akira. "He's not that kind of guy," says Takeuchi. "His only response was, 'So what's next?'"

Make no mistake — Akira is passionate about the research. But, he says, "before I get happy, I must be careful. My greatest fear is an incorrect article".

Akira is a man of few words but many publications. For two years running, he has been the world's 'hottest' researcher, according to citation-based rankings by Thomson Scientific in Philadelphia, Pennsylvania. He won the honour with his prolific studies of the innate immune system, creating an army of knockout mice to help deduce the role of the TLRs. Although many around the world have pushed the field forward, the magnitude of Akira's contribution is undeniable. "He owned that field," says Alan Aderem, an immunologist at the Institute for Systems Biology in Seattle, Washington.

Akira's conservative nature and quiet demeanour seem at odds with a researcher who stands at the forefront of one of the fastest moving, most competitive fields in science. He keeps a lab of just 20–25 people. And his lab members say that he does not pressure them or direct experiments from the top down. He doesn't even work late, leaving for home at 5 p.m. sharp. He shares his mice and his knowledge generously, even ahead of publication.

So what has kept this mild-mannered man at the top of the game? His colleagues credit his ability to choose hard-working people and access to financial resources that he parlayed into a 'knockout-mouse factory'. But they also

K. MAYAMA

recognize his intuition and insight. "If you were digging for oil, he would know where to dig," says Takeuchi, now an associate professor in Akira's lab.

The Japanese government is prospecting. In September, Akira was awarded a 10-year grant to establish a new immunology research centre at Osaka University. One of five projects selected under a new scheme funded by Japan's science ministry, the centre is expected to assume a role of international leadership². Although the scheme's goals are ambitious, Akira's past proves him adaptable.

Back to basics

In 1980, after two years as a physician, Akira joined the lab of Tadamitsu Kishimoto, an eminent immunologist, at Osaka University. Akira planned to work on T-cell and B-cell interactions, but on his very first day, Kishimoto deposited the 28-year-old in the lab of another scientist, Tasuku Honjo. Honjo worked in an adjoining wing, connected by a bridge between the clinical and basic-research groups. "[Kishimoto] told me I had to learn molecular biology," says Akira. "It was a culture shock. I didn't know what a plasmid or a kilobase pair was." But he accepted his charge. Every day for

two years, Akira went to the Kishimoto lab, flipped the wooden name tag to mark his presence, and then headed over the bridge.

The background in molecular biology, and especially genetics, served Akira well, says Kishimoto. "With no computer databases, he sat at his desk and found important sequences. To me, it was just a bunch of letters." Akira seemed to have a knack for the work. "It was like one of those movies about children with special powers," Kishimoto recalls. Aside from a 2-year stint at the University of California, Berkeley, Akira worked at Kishimoto's lab until 1996, getting his name on a major paper every couple of years. The success kept the sometimes tedious work exciting. "It never got depressing," Akira says. But his colleagues definitely noticed his reclusive nature. Kishimoto laughs as he remembers the day a senior professor from a neighbouring lab — the likeable, chatty Tadatsugu Taniguchi — asked him: "Does Akira not like me? He never says hi."

Rather, Akira's work consumed him. In a 1994



study³ on the immune-signalling protein STAT3, he used 6,000 mice in 6 months. By the end of the project, Kishimoto says, the mice dealers were telling him, "There are no mice left in Japan". STAT family proteins relay signals from

extracellular immune regulators such as interferon, interleukin,

and epidermal growth factor to the nucleus of cells, turning on specific response genes. STAT3 was a key to understanding how each signal gets a specific response, and the work progressed quickly. Akira's publication came out a week after one⁴ from a lab at Rockefeller University in New York, run by James Darnell. "I didn't even know we were competing until Kishimoto heard about their work at a conference," Akira says. On the back of his strong publication record, Akira became an independent researcher, first at Hyogo College of Medicine in 1996, then back at Osaka University in 1999, just when the field of innate immunity was starting to take off.

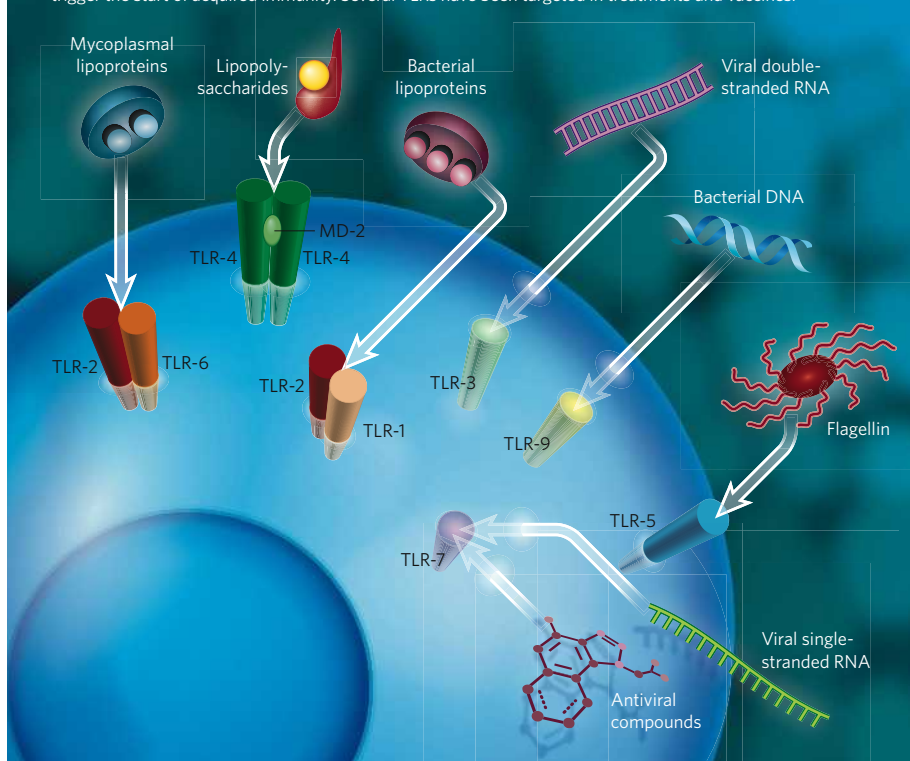
For decades, researchers had focused their attention on acquired immunity, the domain of T-cells and B-cells that rearrange their DNA to target and destroy invaders. This is the system that gives vertebrates the ability to remember infections and to become immunized against them. Innate immunity, on the other hand, is a first line of defence. It had been seen as a ham-fisted relic of evolution, indiscriminately engulfing and digesting incoming microorganisms, mounting a generalized, sometimes powerful inflammatory response, and relaying information about the invading bacterium or virus to the more-sophisticated acquired response.

Toll gates

But innate immunity is much smarter than it had let on, and its genius lies in the TLRs. *Toll*, for which the TLRs were named, was originally identified as a gene involved in the embryonic development of *Drosophila*. Experiments by Jules Hoffmann at the Institute of Molecular and Cellular Biology in Strasbourg, France, in 1996 showed that the gene also helped to protect the flies against fungal infection⁵. In 1997, a human homologue, a TLR, was found to have a function in immunity⁶. And in 1998, Bruce Beutler, a geneticist at the Scripps Research Institute in La Jolla, California, used genetic-mapping techniques to identify a mutation in the *Tlr-4* gene as being responsible for the failure of some strains of mice to produce the often fatal

MIXED MESSAGES

Shizuo Akira's knockout mice helped elucidate the function for seven of the ten human toll-like receptors (TLRs). Different bacterial and viral components trigger the TLRs to elicit an immediate innate immune response and trigger the start of acquired immunity. Several TLRs have been targeted in treatments and vaccines.





Shizuo Akira (left) shares a laugh with his colleagues Yutaro Kumagai (centre) and Osamu Takeuchi.

toxic-shock reaction after they were infected with Gram-negative bacteria⁷. Without the protein to recognize and respond to the presence of the bacteria, the mice didn't have the powerful reaction that would kill them. Coincidentally, as Beutler's article appeared, Akira was preparing his own manuscript showing that purposely knocking out *Tlr-4* produced mice just like Beutler's⁸.

Genetic data showed that mice had 12 TLR genes, 9 of which were very similar to human genes. With expertise in the field and a head start on some of the other knockout mice, Akira helped to pin down the function of seven of these TLRs in short order, showing what kind of invaders they recognized and in some cases how they exerted their effects. The TLRs have a surprising ability to recognize patterns specific to different components of antigens: lipopolysaccharides — huge fat and carbohydrate residues on the surface of Gram-negative bacteria — activate TLR-4, but other TLRs respond to lipoproteins, specific DNA patterns, double-stranded RNA or bacterial flagellin (see graphic). "We thought they might be redundant, but they were all different," says Akira, who calls his discovery, with Takeuchi and Hemmi, of the function of TLR-9, his proudest. Among the long strings of As, Ts, Cs, and Gs in DNA, some parts have dense concentrations of C and G. In mammalian genomes these bases are usually methylated. In bacteria they are not. TLR-9 identifies unmethylated DNA. Such subtle specificity from a seemingly brutish innate response wasn't expected. "People couldn't believe it," he said.

Akira's achievements have gotten his name on some 600 articles. But many of those — roughly half, says Akira off the top of his head

— are collaborations in which his main role was to supply the mouse. The arrangement has certainly helped him become the hottest name in science. But his colleagues add him willingly. According to Aderem, the convention is to credit a researcher who shares a mouse with authorship of one or two articles resulting from work with that mouse. "In his case, they keep putting him on," he says.

Shared success

And there may be reasons for that. Akira's generosity surprised Aderem when he asked for a mouse lacking lipocalin-2, which proved to be a key component in the immune response to bacterial infection⁹. "He just gave it to me," he says, adding that it's extremely rare for people to share their resources before they've gone through them completely. "Absent his mice, the field would not have moved forward," he adds.

Akira's 'factory' has 100–200 knockout mice, Takeuchi estimates. Graduate students each make some 5–10 knockouts, which is an impressive amount. The process, for which the 2007 Nobel Prize in Physiology or Medicine was awarded, requires a series of trial-and-error procedures that can take up to a year and a half per strain to complete. And in the end, the knockout might not even have a clear effect. "If there's no phenotype, it's useless," says Himanshu Kumar, a fourth-year graduate student. Is it fun? "No, not at all. But it's exciting if you're getting good results."

Those in the lab appreciate Akira's advice, but he doesn't push them. "He doesn't have to. The field itself is competitive," says Coban Cevayir, a Turkish postdoctoral student who has worked with Akira for four years.

Akira has now moved beyond TLRs to

address broader aspects of signalling mechanisms within innate immunity. The broader focus set the stage for a collaboration he began last August with Beutler and Hoffmann. The team has a 5-year grant from the US National Institutes of Health to look at host defence against viral infection. "His work is bringing a renaissance to immunology," says Beutler.

And then there's the World Premier International Research Center grant from Japan's science ministry. The grant will provide billions of yen (tens of millions of dollars) to set up new institutions, starting with ¥700 million for to build the Immunology Frontier Research Center, a ten-story building near Akira's lab. The centre's aim is ambitious: "To unveil the whole picture of the dynamic immune system by employing a variety of imaging techniques to visualize the immune cells within live animals." Akira says that previous studies have shown how different molecules function, but not much about how important their response is when a bacteria or virus actually enters the cell. The imaging techniques will show the "spatial and temporal relationships" so "we can see the extent to which different cells contribute to the immune response", he says. "We want to see the immune system as a whole."

The centre's principal investigators will include nine immunologists from Osaka University, including Kishimoto, three others from Japan, including Kyoto University's Shimon Sakaguchi who propelled the field of regulatory T-cells forward, and Fritz Melchers of the Max-Planck Institute for Infection Biology in Berlin, Germany. It will also have a large imaging component, including Osaka University's single-molecule imaging pioneer Toshio Yanagida¹⁰ and six collaborating labs in the United States.

Holding this high-calibre, multi-disciplinary, international project together will be a management challenge for the quiet doyen of immunology. The past couple years of fame and media attention have made Akira much more open and outgoing, Kishimoto observes. But he will probably still keep his comments to a minimum and let his science do the talking. ■

David Cyranoski is Nature's Asia-Pacific correspondent.

1. Hemmi, H. *et al. Nature* **408**, 740–745 (2000).
2. *Nature* **447**, 362–363 (2007).
3. Akira, S. *et al. Cell* **77**, 63–71 (1994).
4. Zhong, Z., Wen, Z. & Darnell, J. E. Jr *Science* **264**, 95–98 (1994).
5. Lemaître, B., Nicolas, E., Michaut, L., Reichhart, J. M. & Hoffmann, J. A. *Cell* **86**, 973–983 (1996).
6. Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A. *Nature* **388**, 394–397 (1997).
7. Poltorak, A. *et al. Science* **282**, 2085–2088 (1998).
8. Hoshino, K. *et al. J. Immunol.* **162**, 3749–3752 (1999).
9. Flo, T. H. *et al. Nature* **432**, 917–921 (2004).
10. *Nature* **408**, 764–766 (2000).

Biofuel: microalgae cut the social and ecological costs

SIR — Much of the discussion over biofuel production has focused on higher plants and the problems associated with their use, such as the loss of ecosystems or the driving up of food prices, as discussed in your Editorial ‘Kill king corn’ and News Feature ‘The little shrub that could — maybe’ (*Nature* **449**, 637 and 652–655; 2007). The use of microalgae, which need not displace food crops, has tended to be overlooked.

Mass algal cultivation has been used to produce pharmaceuticals for more than a decade, so the technology is well developed. Annual yields per unit area are an order of magnitude greater than those for higher plants. Conservative estimates of 30,000–50,000 litres of lipid per hectare per year compare with 1,300–2,400 litres per hectare per year reported for plants with a high oil yield, such as oil palm or jatropha.

Also, biomass production need not compete with food production for either land or water. Both marine and freshwater algae may be used, so either sea water or fresh water can serve as the basis of the culture medium. And as land-based algal culture systems don’t depend on soil fertility, barren land can be used.

Spin-offs in the form of aquaculture would seem feasible, so local communities could benefit from the activity. And in the case of freshwater algae, energy can be recovered from anaerobic digestion of the non-lipid waste: methane produced by this process could supply the needs of the algal cultivation processes and desalination of sea water, or be used for ammonia synthesis.

The cost of biofuel production from algae has been estimated at \$50 or less per barrel (M. E. Huntley and D. G. Redalje *Mit. Adapt. Strat. Glob. Ch.* **12**, 573–608; 2007), making it more than economic at current oil prices, without subsidies.

Peter J. le B. Williams

School of Ocean Sciences, College of Natural Sciences, Bangor University, Menai Bridge, Anglesey LL59 5AB, UK

Biofuel: corn isn’t the king of this growing domain

SIR — We fully agree with your Editorial ‘Kill king corn’ (*Nature* **449**, 637; 2007) that corn (maize) is not a good feedstock for biofuels for a host of environmental, economic and humanitarian reasons.

We also agree that a sustainable biofuels industry needs to rely on non-food crops, such as cellulose from switchgrass or poplar trees, jatropha and possibly corn stover (stalks). Further research and development

is necessary before such feedstocks will become commercially viable.

In the United States, neither the 51 cents per gallon tax allowance given to blenders who mix ethanol (the biofuel derived from corn) into their petrol, nor the 54 cents per gallon tariff on imported ethanol are defensible. The latter serves largely to keep low-cost Brazilian ethanol from sugar cane out of the country.

Unfortunately, the new Farm Bill currently moving through Congress looks almost certain to preserve the farm-subsidy system of direct payments based on a per-bushel price scale for corn and other major crops. Subsidies go mainly to large farm operations, as these produce the most bushels. Corn producers have been among the largest beneficiaries of these subsidies, which encourage overproduction and excessive use of nitrogen fertilizer, the main source of the very potent greenhouse gas nitrous oxide. A new report from the National Academy of Sciences, *Water Implications of Biofuels Production in the United States*, also warns: “If projected future increases in use of corn for ethanol production do occur, the increase in harm to water quality could be considerable.”

The current system of farm subsidies should have been reformed long ago to reward sound environmental practices, set an income cap for those receiving government payments and provide insurance against weather-related disasters.

But rather than killing king corn, we would prefer to say that the emperor — corn-based ethanol — has no clothes.

C. Ford Runge, Benjamin Senauer

Department of Applied Economics, University of Minnesota, St Paul, Minnesota 55108, USA

Research in the wild

SIR — In publishing ‘His daughter’s DNA’ (*Nature* **449**, 772–776; 2007), *Nature* is contributing to an important trend of involving the public in human genetics investigations. Laypeople and advocacy groups are demanding a say in biomedical research by contributing towards its funding and coordination, and even to the development of therapies.

As befits the genomic age, this collective activity is altering interactions among patients, medical professionals and scientists. It is stimulating unlikely alliances among academia, private donors, commerce, patient networks, the media and the state. New forms of solidarity are emerging.

The story of this child and her committed scientist father illustrates the emergence of a new form of social solidarity, founded on genetics. We have become witnesses and participants in a genetics experiment — the girl’s problem and her father’s are made

into our problem. The rest of us need to pay close attention and offer feedback.

Ine Van Hoyweghen, Bart Penders

Department of Health, Ethics and Society, Maastricht University, PO Box 616, 6200 MD, Maastricht, The Netherlands

Kyoto: doing our best is no longer enough

SIR — Gwyn Prins and Steve Rayner are effective in reiterating the Kyoto Protocol’s flaws, in their Commentary ‘Time to ditch Kyoto’ (*Nature* **449**, 973–975; 2007). But their language is unnecessarily negative and they do not deliver a clear, credible alternative.

The Kyoto Protocol is not enough to create the low-emissions transformation in the global economy that is required to tackle the climate problem successfully. But it is already resulting in low-carbon investment and emissions reduction, and is a step towards an effective global treaty. In focusing on the name and arguing that it should be ditched, the authors overlook the potential of a ‘Kyoto phase 2’ that builds on this approach.

Their call for investment in developing technologies as a cornerstone of our response is right, but it doesn’t address the important question of how to achieve it. It takes 20 years for new technologies to get to market — time we do not have. What we need are tools (such as a cost for carbon through market incentives and emissions trading) that facilitate rapid uptake of existing clean technologies. Investment in low-carbon technologies and cap-and-trade measures are two sides of the same coin: to achieve one requires the other. The market is awash with investment funds — good policy is needed to unlock them.

Trying to build a new agreement from the ground up, or hoping for a bottom-up solution to emerge from a mix of domestic responses, simply risks yet another decade of delay, diplomatic wrangling and nationalistic plea bargains while the climate system moves towards catastrophic tipping points.

As Winston Churchill said: “It’s not enough that we do our best; sometimes we have to do what’s required.” That will also hold at the meetings in Bali next month, in Poland next year and in Copenhagen at the 2009 United Nations conference. Pointing out the treaty’s inadequacies is all very well, but the harder and more vital job is building on it to achieve a more effective and adequate one.

Barry W. Brook*, Nick Rowley†, Tim F. Flannery‡

*Research Institute for Climate Change and Sustainability, University of Adelaide, South Australia 5005, Australia

†Kinesis, 255 Riley Street, Sydney, NSW 2000, Australia

‡Copenhagen Climate Council, and Division of Environmental and Life Sciences, Macquarie University, NSW 2109, Australia

BOOKS & ARTS

Flavour and plenty

There may be more to great dishes than a dash of chemistry and a squeeze of lime juice.

Citrus: A History

by Pierre Laszlo

University of Chicago Press: 2007. 262 pp.
\$25

Kitchen Mysteries: Revealing the Science of Cooking

by Hervé This

Columbia University Press: 2007.
232 pp. \$22.95, £13.95

Peter Barham

Many supermarkets offer a year-round array of fresh fruits, vegetables, meats and cheeses. For the first time in history, our eating habits are no longer dictated by seasons or climate. How did we reach this situation of plenty, in which cooking and eating are driven by pleasure as well as hunger? Only a few decades ago, bananas were a luxury in northern countries and strawberries arrived in time for tennis at Wimbledon. *Citrus* and *Kitchen Mysteries* dig into these issues in different and complementary ways.

Pierre Laszlo looks at the widespread availability of citrus fruits as an example of how foodstuffs have been propagated around the world in response to religious, economic and political trends over the centuries. Citrus fruits originated in Asia and were later enjoyed by the Romans and cultivated in southern Europe, having been introduced in about 300 BC by Alexander the Great. The conquering Spanish and Portuguese brought them to the New World. Laszlo highlights technological developments that have contributed to the global spread of citrus fruits — for example, the creation of new strains that can grow in different climates or have increased resistance to diseases or variations in flavour. He describes advances in juice extraction, packaging, storage and transport.

Orange juice as a mass commodity in the United States has its origins in a chance recommendation to an up-and-coming copywriter, Claude Hopkins, by a press baron. Hopkins came up with the timely slogan 'Drink an orange' in 1908, when growers facing a glut of oranges in California were about to start destroying trees. Instead, the excess fruit was turned into juice and sold. The rest, as they say, is history.

Citrus provides a colourful background of the literature, poetry and art associated with citrus fruits, as well as their pharmaceutical



Gastronomy alfresco: does food really taste better outdoors?

effects. Apparently, an ingredient of grapefruit juice deactivates an enzyme in the small intestine that destroys some medications before they can enter the bloodstream. Alternatively, the citrus component boosts the activity of certain drugs, such as sildenafil (better known as Viagra) and inhibitors of HIV-1 proteases.

Rapid development of other crops across the world has led us to the present cornucopia of readily available raw ingredients. How can we transform this bounty into great dishes full of thrilling flavours and textures? Hervé This uses the kitchen as a laboratory to meet this challenge and to analyse the scientific secrets of culinary delights.

In *Kitchen Mysteries*, he describes the scientifically important steps in cooking processes and how they could be improved, revealing tricks of chemistry and physics to prevent soufflés from collapsing and mayonnaise from breaking up. This molecular gastronomy is garnished with the author's own rich philosophy of food and flavour.

The science of taste and flavour is still young. We understand tongue physiology, but still argue about how many taste sensations our tongues can discriminate — there are at least five (salt, sour, sweet, bitter and umami). The olfactory sense is even more controversial — it is unclear how many types of sensor we have, how they work and to what extent we can learn to smell new aromas.

We do have a rough picture of the signals sent to the brain by sensors in the mouth and

nose. But determining the flavour of a dish and deciding whether we like it is complex: we do not simply sum the signals arriving at the brain to 'calculate' the taste. Flavour, and the degree to which we enjoy it, seems to be a mental construction, relying on memories of past experiences as well as on incoming sensations. Everything about a meal — the colour of the plate, the music in the background and even the people eating it with us — can affect our perception of its flavour.

These books should help any experimental scientist to become a better cook, although *Kitchen Mysteries* sometimes optimistically assumes that we all know how to make a béchamel sauce or why highly methoxylated pectins are crucial for jam-making. In *Citrus*, Laszlo applies his expertise as a chemist to the understanding of citrus fruits. His succinct explanations of the Maillard and caramelization reactions when describing how to make marmalade, and his notes on foams, phase separation, and the volatility of aromas when explaining how to make a sabayon, justify the cost of the book on their own.

Both books have one drawback — neither has an index. This detail could avert a catastrophe for your guests by helping you quickly to locate, for example, advice on what to do if your Béarnaise sauce fails to thicken (remove from heat, add a little water, then beat very hard). ■ Peter Barham is a professor in the H. H. Wills Physics Laboratory, University of Bristol, Bristol BS8 1TL, UK.

Quests of a theoretical astronomer

Practical Mystic: Religion, Science, and A. S. Eddington

by Matthew Stanley

University of Chicago Press: 2007.

320 pp. \$37.50

Owen Gingerich

In this extraordinary book about the life of the distinguished English astrophysicist Sir Arthur Eddington, Matthew Stanley examines the entangled roles of science and religion in his work. *Practical Mystic* is not a biography in the ordinary sense — readers will look hard to find dates for Eddington's birth and death (1882–1944), and much else has been omitted. But included in rich detail are Eddington's Quaker milieu and the tensions he faced at the tribunal as a conscientious objector during the First World War, his astrophysical research, his pioneering fascination with Einstein's relativity, and his role as a major popular writer on astronomy and the philosophy of science.

I can recall only one other book that attempts to build a convincing biographical interconnection between a religious ethos and scientific achievements — Job Kozhamthadam's *The Discovery of Kepler's Laws: The Interaction of Science, Philosophy, and Religion* (1994). Trained as a theologian, Kepler saw God's design throughout the cosmos. Whether that drove him to search for the physical (rather than the traditional geometrical) understanding that led him to his three laws seems inconclusive. In Stanley's analysis of Eddington, there is no doubt of the compatibility and mutual influence of science and religion.

To analyse the relationship between science and society (including religion), Stanley examines the bridging function of what he calls "valence values". Like the bonding ring of electrons, these values facilitate the interaction between science and culture. Through the lens of these values, Stanley uses Eddington as a test case for exploring the interaction of science and religion in Britain in the first half of the twentieth century.

Unlike the natural theologians of the previous century, Eddington did not seek a harmonization between science and religion. He saw both as processes of seeking. As he reminded his audience at the British Association for the Advancement of Science, "A knowledge of nature is the great end of our work; but, if we cannot attain that, there is at least the struggle after knowledge, which is perhaps no less a thing." Eddington could have said the same of his religion.

When he approached fundamental questions of astrophysics, Eddington did not try first to establish basic laws from which conclusions could be deduced (as Newton had done and as his rival James Jeans insisted on doing), rather he built a web of approximations whose

results could be compared with nature. Instead of first asking specifically what could power the Sun, he worked through possible structures to establish the likelihood of an extremely hot, temperature-dependent core as the seat of the Sun's energy source. (He suggested his critics should find a hotter place.)

Out of the same search procedures came his famous law linking the luminosity of stars with their mass. Stanley describes the astrophysics with considerable skill and using essentially no mathematics. It must be a little perplexing to the non-specialist though when the expression " κ " appears with no definition, nor does Stanley explain that " κ " stands for opacity.

Eddington introduced Einstein's general theory of relativity to the English-speaking world. Stanley tells this story well, arguing that Eddington's Quaker faith and his pacifism motivated his desire to test the work of a German scientist at a time when, after the First World War, Germany and its scientists were loathed by the Anglophone community.

The results of the 1919 eclipse expedition, which showed stars near the Sun slightly displaced by the curvature of space associated with its mass, made Einstein famous overnight and thrust Eddington into the limelight too.

A story — now almost an urban legend

— has gained currency that Eddington was so convinced of the accuracy of Einstein's prediction of the bending of starlight by the Sun, that he fudged the treatment of marginal eclipse plates to obtain the desired results. Here Stanley

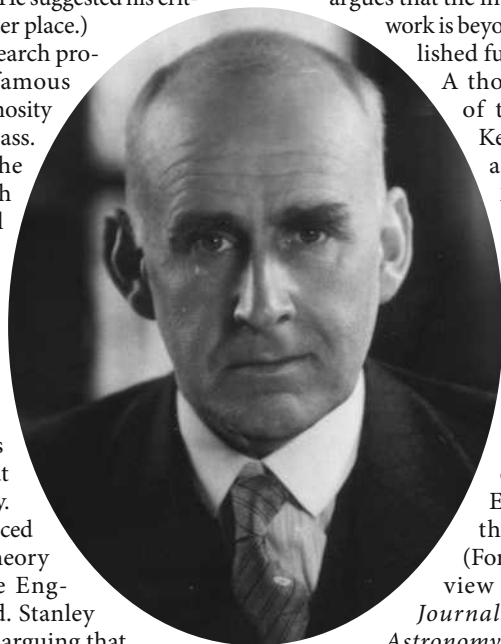
argues that the integrity of Eddington's work is beyond doubt (he has published full details elsewhere).

A thorough examination of the case by Daniel Kennefick (preprint at arXiv:0709.0685 v2; 2007) substantiates Stanley's claim.

Among the omissions in Stanley's biographical study is the put-down of S. Chandrasekhar's calculations of the ultimate collapse of highly massive objects, something Eddington found aesthetically distasteful. (For a perceptive, neutral view of this episode see *Journal for the History of Astronomy*.) Perhaps this also connects with Eddington's aversion to the idea of the Universe starting at a finite time past, something that may have

smacked too much of the ultimate miracle for his religious outlook and a dead end for his philosophy of seeking. Stanley's *Practical Mystic* is not a biography but a biographical study — a fascinating one. ■

Owen Gingerich is professor at the Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA and author of *God's Universe*.



Eddington saw science and religion as processes of seeking.

Time deconstructed

The New Time Travelers: A Journey to the Frontiers of Physics

by David Toomey

W. W. Norton: 2007. 320 pp.

£15.99, \$25.95

Lawrence M. Krauss

There are innumerable books on time travel, including the masterful *Black Holes and Time Warps* by Kip Thorne, one of the key players in a serious scientific investigation of the possibility of time travel. The subject has a chequered history, peppered with false claims and hyperbole. The title *The New Time Travelers* had me worried that the book belonged to

'People magazine popular science' — the genre that propagates the false notion that somehow researchers are more interesting than research. Happily, David Toomey, professor at the University of Massachusetts, Amherst, has produced an honest, intelligent and largely accessible work of impressive scholarship on a difficult subject. Most important, he provides a rare glimpse into the day-to-day practice of science, in which the right direction is never clear and false starts abound.

After some historical, scientific and philosophical scene-setting, the book focuses on the surge of attention, which began almost two decades ago, on the question of whether

general relativity allows for the possible and practical realization of time travel. This heightened interest followed the 1988 paper published in *Physical Review Letters* by Thorne and his colleagues on the physics of wormholes, shortcuts through space and perhaps time. The paper was inspired by a question Carl Sagan posed to Thorne when writing his science-fiction novel *Contact*.

The discussion of time travel in a mainstream science publication attracted media attention and piqued interest in the scientific community. It also encouraged a lot of fanciful speculation by scientists about ideas on the very edges of respectability. And it motivated the writing of numerous popular books.

Since the mid-1990s, most concrete results regarding the possibility of time travel have been negative. The few outstanding issues ended up mired in quantum gravity, which remains ill-defined and elusive. As a result, interest in this issue in the physics community has subsided somewhat.

Toomey remains enamoured with the philosophical and literary possibilities of time travel, so is less critical of some dubious proposals than he might be. He nevertheless provides a largely accurate rendering of time-travel science with all its twists and turns, errors and dead ends. This is sometimes frustrating: it is tough to labour through detailed discus-



The physics of wormholes: are there shortcuts through time?

sions of complex ideas that Toomey eventually correctly reveals to have proved fruitless. But it is also refreshing: popular books too often present research as a logical narrative that bears little relation.

Errors, as might be expected from someone writing outside his field, have crept in. Early

discussions of physics, including special and general relativity, are quite clear. Later explanations of modern topics, such as inflation, are more uneven. A grating mistake is Toomey's statement that the Planck length is a million times smaller than the size of a proton. It is actually closer to a million million million times smaller. We regularly investigate regions a million times smaller than a proton with large-particle accelerators: it is the vast chasm between even these minute sizes and the Planck length that makes probing quantum gravity so difficult.

Aside from such gaffs, *The New Time Travelers* is sound and informed. If readers leave this book feeling unsatisfied, Toomey may have done them a service. He captures the nature of the scientific process — that one never knows where fundamental research will lead, that most of it doesn't result in ground-breaking developments, and that even negative results can prove enlightening. It is sometimes difficult for us to be honest about these facets when popularizing our work. That an 'outsider' has captured this is impressive, and useful. ■

Lawrence M. Krauss is director of the Center for Education and Research in Cosmology and Astrophysics at Case Western Reserve University, Cleveland, Ohio 44106, USA. He is author of *Hiding in the Mirror: The Quest for Alternate Realities, from Plato to String Theory*.

Scientists on film

Hollywood Science: Movies, Science, and the End of the World

by Sidney Perkowitz

Columbia University Press: 2007. 256 pp. \$24.95

Emma Marris

Science-fiction films are often called to account by scientists — for their inaccuracies and for skipping over the years of grueling effort behind discoveries. But film is not particularly interested in reality. It is obsessed with unreality, which, after all, is why we go to the movies. We get reality at home. *Hollywood Science* reveals, perhaps inadvertently, what scientists owe to film: a kindling of interest in scientific concepts that shadows the audience as they leave the cinema.

In film, science and scientists act as a magic key to the unknown. They take us into realms where people and animals mingle in the same individual, where heads live apart from bodies, where aliens and robots and clones wear shiny jumpsuits and talk to their watches and make love. In the cheesy 1994 film *Stargate*, a hammy James Spader plays an Egyptologist who saunters into a military industrial complex and

does his science thing, which enables guys with guns to walk through a door into another world. In films involving an atomic bomb, scientists turn a key that leads to a bleak but fascinating future. Science provides new modes of transportation and new destinations: time machines to the future and rockets to the stars.

Yes, researchers may cringe at some of cinema's favourite lab-coated characters — the effeminate egghead who wants to negotiate with murderous aliens, or those government jerks in hazmat suits who wanted to carve up E.T. But the important thing, perhaps, is that science and scientists continue to be used to expand the world of film in imaginative ways. These themes reveal that, despite its anti-science spasms, the culture at large believes that science is a means to go places, see things and know more, for good or ill.

Over the years, scientists have complained that they are not in fact megalomaniac geniuses with delusions of godhead that prompt their hair to stand on end, like Victor Frankenstein, or Dr Rotwang in *Metropolis*. This earnest refutation does not do much to dispel another scientific stereotype: that of the humourless pedants immune to the communal pleasures



Dennis Quaid, with stern expression, as a scientist in *The Day after Tomorrow* in 2004.

of humanity. Dennis Quaid plays one such caricature in the climate-change-themed *The Day After Tomorrow*, in which he maintains a grimly furrowed brow, perhaps to detract from the scientific howlers in the film.

Sidney Perkowitz, a physicist at Emory University in the United States, keeps up the tradition of examining the stock cinematic geeks. He says that the stereotypes are not entirely without basis, observing that, on average, one-third of the men at scientific conferences have facial hair and half of both sexes wear glasses. Scientists can be arrogant, he maintains, and “real scientists aren’t usually the brightest sparks in the room — often interestingly quirky, yes, but not necessarily loads of fun.”

Perkowitz’s book is an affectionate examination of dozens of science-fiction films. If you don’t want endings spoiled, avoid this book.

Detailed plot summaries alternate with short, layman-level explanations of the research behind such staples as asteroids hitting Earth, computers taking over, atomic holocausts and alien invasions.

It is not clear whether Perkowitz has written this book for teenagers or whether his exuberance transports him into a state of child-like rapture, but what comes across is a slightly naïve enthusiasm. This may speak to the difficulty of switching from science to cultural criticism.

He saves his scorn for just a few movies. One, *The Core* (a 2003 flick in which the Earth’s core stops rotating and has to be restarted by a team drilling down to the Earth’s centre to set off nuclear explosions), seems to have been made deliberately to annoy scientists. Another, *What the #\$! Do We (K)now!*, is actually a promo-

tional film for an affluent US cult, led by a woman in Tacoma, Washington, who ‘channels’ a 35,000-year-old warrior. Pretty much everything else, Perkowitz likes. Evil scientists, wimpy scientists, boring scientists — as long as the ideas are entertaining, he’s on board. And that is because, like most scientists and most popcorn-munching proto-scientists, he is curious and enthusiastic.

It is these very qualities that motivate both the science-fiction aficionado and the scientist, although the latter also needs patience, stamina and rigour. And that’s why, for all their flaws (yes, the fact that the Hulk is larger than Dr Bruce Banner does violate the first law of thermodynamics), science-fiction films — good, bad or indifferent — are important. ■

Emma Marris is a retained correspondent at *Nature*.

Dishing the dirt on hygiene

The Dirt on Clean: An Unsanitized History by Katherine Ashenburg

North Point Press: 2007. 368 pp. \$24
Knopf Canada: 2007. 368 pp. Can\$35

Virginia Smith

Nothing on cleanliness for 50 years and then two books come along at once — mine and Katherine Ashenburg’s *The Dirt on Clean*. If anyone ever wanted to know about the biological, psychological, social and technical history of cleansing, it’s all out there now. The history of hygiene is a new field that may yet influence public policy.

Ashenburg’s neat distillation of historical anecdotes and graphic biographical details of personal hygiene is more focused on baths and bathing than on the principles of hygiene *per se*, in the mould set by Lawrence Wright’s international best-seller *Clean and Decent* (1960). She does not attempt to cover medical therapeutics, natural philosophy, legislation, religion, sociology, grooming, or indeed anything before the Greeks. That gives her more space to investigate her key proposition: whether or not Europeans and Americans in the past were genuinely dirty and smelly.

She looks at bathing habits from the classical world to the twentieth century, borrowing at length from a wide range of letters, diaries and travel writings, combined with a good array of secondary historical sources. This approach allows the reader to relish all those salacious, deliciously taboo details of dirt, filth and personal body odour.

Ashenburg uses the psychology of dirt to provide insight into the later germ-inspired cleanliness crusade and the more recent, advertising-led campaign for personal hygiene. The fear of dirt is a psychological threshold that can be either raised or lowered, and Europeans were obviously fairly relaxed about it until the arrival

of microbiological ‘germs’. Even that fear faded during the past century — until the recent spate of antibacterial advertising. Ashenburg ends by addressing the fearful excesses of modern North American hygiene. These have become a burning issue in the light of the immunological theory relating hygiene to the soaring increase in child allergies in the Western world.

We are at the start of a long-overdue review



Fifteenth-century bathers in Baden, Switzerland, compromising their ablutions with decorative hats.

of nineteenth-century propaganda, which assumed (without much evidence) that a Dark Age of dirt began after the fall of Rome and was put right only by modern sanitation. Demographers and epidemiologists may still think that this was the case, but historians have been digging up some colourful cultural details that smudge this neat picture. They have confirmed that classical and mediaeval Europeans liked to bathe, and that plague and syphilis closed the

public bath houses and put people off all but the most essential ablutions for a while during the seventeenth century.

Bathing habits had to be relearned during the eighteenth and nineteenth centuries. It was the long modern economic boom that allowed most industrialized countries to renew their civic infrastructure with water supplies and indoor plumbing. As recently as 50 years ago, only one in ten French homes and just over half of British and US homes had a plumbed bathroom. These figures obscure the twentieth-century divide between urban and rural living conditions: most urban homes had a bath or shower by the 1950s, whereas rural areas did not catch up for another 10 to 20 years.

There are other deep-rooted themes emerging from the new history of hygiene that Ashenburg’s bathing history cannot take into account, namely politics and medicine. Hygiene was the political force behind the democrat and the religious dissenter. It has given us enlightened sanitationism and much-valued civic amenities such as parks, libraries and swimming pools; it contributed to eugenics race theory and to the contemporary consumer-based, science-led ‘green’ politics of pollution and climate change.

The history of health education and the development of a ‘health regimen’ underpin today’s ideas of healthy behaviour, now promoted by government risk-awareness legislation, food-and-exercise campaigns, and regulation of ‘work–life balance’. Another underrated therapeutic theme is beauty care, also booming at the moment. Personal grooming turns out to be something that we brought with us from the Neolithic, and is a form of nurture that, arguably, we can ill dispense with. If, like chimps, we spent many hours a day grooming back then, small wonder we spend time on it still. ■

Virginia Smith is an honorary research fellow at the London School of Hygiene and Tropical Medicine, London N19 5JS. She is the author of *Clean: A History of Personal Hygiene and Purity*.

Accommodating dissent

Providing cures for health problems isn't enough, if people's personal or cultural beliefs clash with the scientific approach. Policy-makers must recognize and engage with these objections.

Melissa Leach

Dismissing public and political concerns about health interventions as unscientific, irrational or misled fails to do justice to the different perspectives in play. It widens gulfs between those promoting the technologies and those who could benefit from them. It is why we see backlashes to even the best-intentioned initiatives. These backlashes signal the need for new forms of dialogue to enable scientists, policy-makers and citizens to navigate the multi-layered politics emerging around science and technology.

For example, in the Gambia, programmes of antiretroviral treatments for HIV and AIDS have been disrupted since January 2007 when its president, Yahya Jammeh, announced a 'cure' for AIDS based on Koranic incantations and herbs. As people living with HIV have come forward to join the president's programme amid heavy state publicity and security, biomedical critics have found it difficult to be heard, and to dispute the controversial laboratory blood tests that Jammeh claims prove the efficacy of his cure.

In Nigeria, the Global Polio Eradication Initiative was derailed in 2003 when three northern states rejected the oral polio vaccine. Datti Ahmed, a doctor and president of Nigeria's Supreme Council for Sharia Law, told journalists: "We believe that modern-day Hitlers have deliberately adulterated the oral polio vaccine with antifertility drugs and contaminated it with certain viruses which are known to cause HIV and AIDS." Although vaccination was allowed to resume in 2004, worry in northern communities still besets the campaign. Polio is now on the rise in the area, and is spreading across Africa.

Such anxieties about and rejection of science, its technologies and safety, are not exotic phenomena. In Britain, parental concern about the measles, mumps and rubella (MMR) vaccination has led to an alarming rise in the rate of measles among incompletely vaccinated populations since the late 1990s.

Scientists, policy-makers and professionals often respond to such controversies by urging that 'sound science' must rise

above politics. This unwittingly casts campaigners and the public as ignorant, irrational, mistrustful or simply misled by the media. Yet such images are themselves simplistic and misleading. They overlook the positive reasons why people question or seek alternatives to science, grounded in their own knowledge and experience. Equally misleading are policy stereotypes proposing that trust and rationality are breaking down in Europe, whereas in Africa they are not yet achieved. In both cases, health policy and professional organizations say that the solution is simply better communication of the science and risks involved.



Better communication is neither simple nor is it the whole solution. Instead we must acknowledge the deep and textured politics of knowledge that social scientists are working to describe. In health care, for example, an individual's understanding of their body is integrated with their social and political experience, and this integration can be seen from various perspectives including those of the patient, the clinician and the policy-maker.

Thus in Britain, the perspectives that questioned MMR integrated ideas of personalized immunity and vaccine-vulnerability (as in the view that 'MMR may be safe but not for my child'), personalized ethics in social affairs and a liberal politics of choice. These perspectives clashed with those linking epidemiology and population-level risk with social and political arguments for mass vaccination.

In West Africa, many people appreciate vaccination for keeping them healthy (often expressing this as 'strength' of body and blood) and routine primary health-care

systems for building strength in local communities. The oral polio-vaccine campaign in Nigeria, however, was set up by outsiders to target one particular disease rapidly over a large area. It therefore clashed with the perspectives in West African communities and in so doing, played into prevailing political tensions. From a local perspective, the campaign was not contributing to local bodily and political strength but threatening it. This is partly because polio, with its slower progression and more varied presentation, is not seen as a priority compared with the diseases such as dysentery, malaria and dengue which people routinely experience as killing children under five in their communities. The 'drops for our children's mouths' were thus culturally interpreted as having other effects, such as on fertility.

The campaign's mass of resources and political attention failed to boost the locally valued primary health-care systems that were in a state of near collapse. Local authorities and people alike interpreted the campaign in wider political terms, as examples of international, central and southern Nigerian political interests seeking to undermine further the strength of the northern states.

A more balanced appreciation of such diverse perspectives is vital. It could be assisted by effective collaboration among social scientists, natural scientists and decision-makers in designing and planning health programmes, as well as responding to controversies. International agencies and national health services could promote and facilitate this. New forums facilitated by independent bodies — such as non-governmental organizations — will also be needed to bring such agencies together in deliberation with scientists, local authorities and citizens' groups. This will help ensure that investments in science and technology — whether for downstream applications or for research and innovations — are firmly enmeshed with inclusive debate about the social and political values they serve.

Melissa Leach is director of the ESRC STEPS Centre, Institute of Development Studies, University of Sussex, Brighton, East Sussex BN1 9RE, UK.

For more essays and information see <http://nature.com/nature/focus/scipol/index.html>.

D. PARKINS

SCIENCE & POLITICS

STEM CELLS

Primates join the club

Ian Wilmut and Jane Taylor

Researchers have achieved the testing goal of generating embryonic stem cells from the cells of an adult primate. The procedure used could provide insights into a variety of diseases, if it can be applied in humans.

Generating patient-specific stem cells holds great promise for medical therapy, drug discovery and basic research. But, quite apart from the associated ethical issues, there are many technical hurdles to achieving this goal for humans. For many species, both adult and fetal cells have been used to generate genetically identical offspring, the most famous of which is probably Dolly the sheep¹. This has been achieved by a process known as somatic-cell nuclear transfer (SCNT), which involves replacing the nuclear genetic material of an unfertilized egg with that of a somatic (non-germ) cell. In each case, species-specific modifications to this basic protocol must be made. Yet despite considerable efforts, it has remained impossible to generate primate-derived embryonic stem-cell lines or to clone a primate. Byrne *et al.*² (page 497 of this issue) now describe the generation of stem-cell lines from embryos of rhesus macaques using a modified SCNT protocol*.

*This News & Views article and the paper concerned² were published online on 14 November 2007.

Byrne and colleagues' earlier work³ had shown that, in contrast to the situation in other mammals, once the nuclear material of an egg (oocyte) is replaced with that of a primate somatic cell, the oocyte cannot efficiently undergo its essential remodelling to an embryonic state because of delayed and inefficient removal of a protein called lamin A/C. This protein is a component of the nuclear skeleton, and its removal requires a protein complex known as maturation-promoting factor (MPF). The authors hypothesized³ that inappropriate handling of the oocyte might cause premature decay of MPF activity, leading to the persistence of lamin A/C. To test this, they inhibited MPF degradation and found that, indeed, lamin A/C was more efficiently removed. But this treatment resulted in other deleterious effects at later stages of embryonic development, which outweighed its benefits.

Byrne *et al.*² have since made several modifications to their SCNT protocol. For example, in conventional SCNT, the oocytes are exposed to a Hoechst stain and/or ultraviolet light

to locate and remove their chromosomes. Byrne and colleagues postulated that such treatment might be responsible for preventing the fertilized primate oocyte from reaching the multicellular, blastocyst stage of embryonic development at a desirable rate. They also thought that it might damage mitochondrial DNA, hindering the function of this energy-generating organelle during early embryonic development. To avoid such potential hazards, the authors adopted an alternative method of chromosome detection: they used the Oosight imaging system to visualize chromosomes using polarized light.

Another measure taken was to increase the level of MPF by removing calcium and magnesium from the media in which the oocytes, and subsequently the embryos, were handled. Although calcium is essential for the oocyte's response to the activating effect of sperm during fertilization⁴, environmental factors can also activate oocytes spontaneously. Thus, the removal of calcium — and maybe magnesium — during SCNT procedures reduces

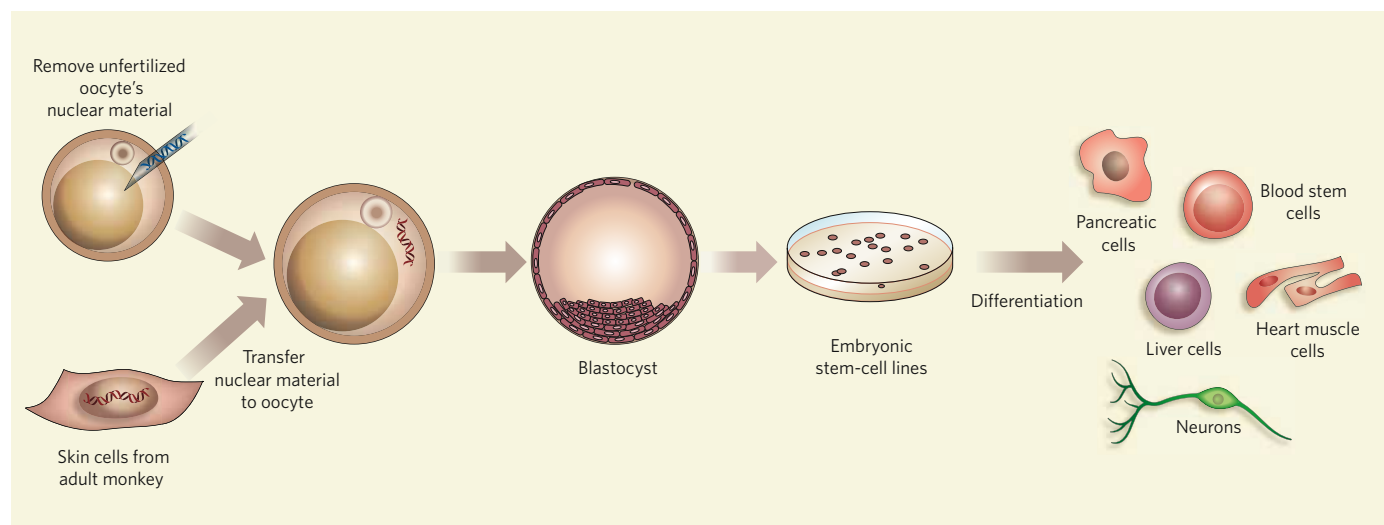


Figure 1 | The technique of somatic-cell nuclear transfer (SCNT). In much the same way as women undergoing *in vitro* fertilization procedures are treated to make them super-ovulate, Byrne *et al.*² treated female rhesus monkeys with hormones to induce the shedding of extra eggs. After recovering these cells, the authors removed the cells' nuclear genetic material. Meanwhile, they obtained skin cells from an adult male monkey, allowed these to multiply in culture, and then treated them

to halt their progress through the cell cycle once they had entered the resting phase known as G0. Next, the authors extracted the nuclear genetic material from the skin cells and introduced it by electric pulses into the nucleus-free eggs. The fused cells were allowed to reach the blastocyst stage of embryonic development before embryonic stem cells were derived from them. Such cells have the potential to differentiate into different cell types.

the incidence of spontaneous oocyte activation. This allows a set of uniform oocytes with high levels of MPF to be generated, which can then be activated in a controlled manner. Byrne and colleagues' success² in generating primate-derived embryonic stem-cell lines followed the introduction of these changes to the SCNT protocol (Fig. 1). However, further work is required to unravel the mechanisms underlying this achievement.

An entirely different approach to SCNT can also lead to embryonic stem-cell lines. This procedure involves reprogramming adult cells directly by inducing the expression of just four gene transcription factors in the cells. Using this method, researchers^{5,6} have reprogrammed skin cells of adult mice to generate a few cells that have many of the characteristics of embryonic stem cells.

There is so far no sign that this approach could be effective in human cells. But even if it could be applied to generating patient-specific cells, there are other limitations on its use in humans. For example, viral vectors were used to introduce the genes encoding the four transcription factors into the genome of the mice. Because of the potential risks associated with using viral vectors, this procedure would be unacceptable for use in treating humans. Moreover, many of the embryonic stem-cell-like cells generated in this way eventually developed into tumours, probably because misregulation of one of the four introduced genes — *Myc* — can lead to cancer. But a modified approach to direct reprogramming that does not involve cancer-causing genes or the use of viruses is likely to be the ultimate method of choice for producing human stem cells.

What are the implications of Byrne and colleagues' findings² for applications in humans? When considering the potential of stem cells derived from adult cells, great emphasis is often placed on the fact that derivatives of such cells, if returned to a patient suffering from a degenerative disease, would not be rejected by their immune system. Realistically, a careful examination of resources and the time required to produce differentiated cells for treatment purposes suggests that large-scale use of stem cells would be impractical.

In our haste to use patient-specific cells in therapy, however, we tend to overlook the fact that they are of great value for basic research and drug discovery. For example, such cells could provide new ways to study inherited diseases. If the diseased tissue cannot easily be recovered from the patient, production of patient-specific cells is the only potential means of obtaining cells with the condition. Such cells could be used to identify causative mutations. Or they could be compared with their healthy counterparts to identify the molecules and molecular mechanisms underlying the disease symptoms. This information could then form the basis for high-throughput screens to identify small-molecule drugs that could prevent such disease-associated changes at a molecular

level. Ultimately, this approach might lead to treatments for neurodegenerative diseases, some cancers and psychiatric disorders. ■

Ian Wilmut and Jane Taylor are at the Centre for Regenerative Medicine, Chancellor's Building, University of Edinburgh, 49 Little France Crescent, Edinburgh EH16 4SB, UK.
e-mails: ian.wilmut@ed.ac.uk;
jane.taylor@ed.ac.uk

1. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. S. *Nature* **385**, 810–813 (1997).
2. Byrne, J. A. et al. *Nature* **450**, 497–502 (2007).
3. Mitalipov, S. M. et al. *Hum. Reprod.* **22**, 2232–2242 (2007).
4. Whitaker, M. *Physiol. Rev.* **86**, 25–88 (2006).
5. Okita, K., Ichisaka, T. & Yamanaka, S. *Nature* **448**, 313–317 (2007).
6. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).

MATERIALS SCIENCE

Purity rolled up in a tube

Fotios Papadimitrakopoulos and Sang-Yong Ju

Before carbon nanotubes can fulfil their potential in device applications, better ways must be found to produce pure samples of them. A promising approach involves wrapping them up in a shell of polymer.

Formed simply by rolling up a two-dimensional sheet of graphite (graphene), single-walled carbon nanotubes — SWNTs — are wonder materials of modern materials science. They are phenomenally strong and stiff, and, unusually, are excellent conductors of heat along the tube's axis, yet good thermal insulators across it. But it is their electrical characteristics that excite the most interest: depending on the precise way they are rolled up, they can be either semiconductors or fully conducting metals.

And there's the rub. SWNTs can generally be fabricated only in batches that vary widely both in the diameter of the individual tubes and in the orientation of their graphene lattice relative to the tube axis — the property known as chirality (Fig. 1). Separating out different conformations to produce a 'pure' nanotube sample is a thorny problem, but one that must be solved if nanotubes are to fulfil their electrifying potential.

Writing in *Nature Nanotechnology*, Nish et al.¹ show how the selection of a specific nanotube chirality can be greatly improved by using a polyfluorene-based conjugated polymer as a wrapping agent. Chen et al.², writing in *Nano Letters*, verify that finding, and achieve further enrichment by using related conjugated polymers to preferentially wrap nanotubes with less variation in their mean diameter. Quite apart from the importance of these techniques for nanotube separation, this work^{1,2} allows a peek at future polymer–nanotube architectures, which could be poised to revolutionize photovoltaic, sensor and flexible-display technologies in their own right.

So how does this chirality enrichment work? Just like graphite³, SWNTs have a smooth surface, and interactions with this surface can be used to organize many chemical species. The interaction between SWNTs and species containing aromatic groups, such as polyfluorene, has been argued⁴ to proceed first by interactions

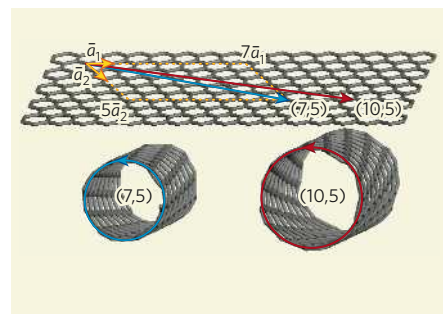


Figure 1 | Roll up. The diameter and chirality (orientation of the graphene lattice with reference to the tube axis) of a single-walled carbon nanotube can be described by a two-dimensional 'rolling vector' based on unit vectors (\vec{a}_1 , \vec{a}_2) along two axes of the unrolled graphene lattice. The conformation is determined by joining the origin and end point of the rolling vector.

between the electrons of the graphene lattice and the aromatic groups, followed by a stage of close packing that depends strongly on the underlying chirality of the SWNTs.

Polyfluorene's chemical structure falls into the basic category of 'hairy rod' conjugated polymers (Fig. 2a). The hairs take the form of short side-chains, formed of several CH_2 groups and one CH_3 group, which are attached to a fluorene backbone, the rod. Additional aromatic groups inserted between the fluorene groups make the polymer repeat-unit longer. The polyfluorene chains are organized into planes such that one side-chain is aligned between adjacent chains and one extends upwards to interact with the solvent and impart solubility to the structure (Fig. 2b).

The overall result is a smooth surface that encircles the equally smooth surface of the SWNTs and thus can guide the wrapping process (Fig. 2c). Crucially, however, these two smooth surfaces have fine corrugations. As the latest papers describe^{1,2}, when the

polymer sheet is rigid, only selected nanotube chiralities match the 'corrugations' of the polymer sheet, kick-starting the wrapping process and facilitating chirality enrichment.

Nish *et al.*¹ also show that, depending on the nanotube diameter required (between 0.5 and 1.4 nanometres), two, three or four polyfluorene chains are needed to wrap up a SWNT completely. Both sets of authors find^{1,2} that selective chirality enrichment is influenced by the length of the side-chains and the precise structure of the polymeric repeat. These features give the polyfluorene sheet rigidity, leading to better recognition of the nanotube corrugation. Chen *et al.*² also demonstrate that the ratio of polymer to nanotubes is a decisive factor for chiral selectivity, indicating that the formation of such superstructures is dynamic rather than static.

The precise organization and directionality of aromatic surfactants in processes for sorting SWNTs has been a focus of intense theoretical study⁵. Recent experiments⁶ have also shown that a chiral biporphyrin chemical group can be used to separate certain SWNTs into optically active nanotube fractions, depending on whether the tube is rolled up in a right-handed or a left-handed way. Such separation

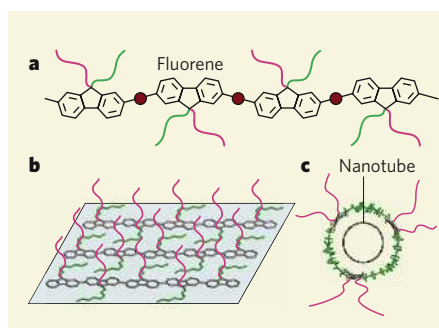


Figure 2 | Hairy rod as guiding hand.

a, Polyfluorene's chemical structure consists of a fluorene backbone connected by additional aromatic groups (red circles). As the authors show^{1,2}, both this conformation and the side-chains of the structure are crucial to rigidify (green side-chain) the polyfluorene structure and also to make it soluble (pink side-chain). **b, c**, This allows several polyfluorene chains to bind together and form a rigid sheet that selectively interacts (**c**) with specific nanotube diameters and chiralities.

adds credence to the idea of an assembly process in which aromatic adsorbates seem to 'communicate' with the underlying chiral side-wall corrugation.

So what's next? Clearly, the work of Nish *et al.* and Chen *et al.* provides a starting point for future investigations of specific polymer–nanotube interactions that take into account the atomic characteristics of both structures. If talented chemists, physicists and surface scientists club together, the result could be nanotubes that not only are enriched in a specific chirality, but also interact with polymeric substrates to produce nanocomposites with specific desired properties. A whole new world of devices is poised to take advantage of what might emerge from such efforts.

Fotios Papadimitrakopoulos and Sang-Yong Ju are in the Nanomaterials Optoelectronics Laboratory, Polymer Program, Department of Chemistry, Institute of Materials Science, University of Connecticut, Storrs, Connecticut 06269-3136, USA.
e-mail: papadim@mail.ims.uconn.edu

1. Nish, A., Hwang, J.-Y., Doig, J. & Nicholas, R. J. *Nature Nanotech.* **2**, 640–646 (2007).
2. Chen, F., Wang, B., Chen, Y. & Li, L.-J. *Nano. Lett.* **7**, 3013–3017 (2007).
3. Tao, F. & Bernasek, S. L. *Chem. Rev.* **107**, 1408–1453 (2007).
4. Zheng, M. *et al.* *Nature Mater.* **2**, 338–342 (2003).
5. Lu, J. *et al.* *J. Am. Chem. Soc.* **128**, 5114–5118 (2006).
6. Peng, X. *et al.* *Nature Nanotech.* **2**, 361–365 (2007).

MICROBIOLOGY

Woodworker's digest

Andreas Brune

Termites digest wood with the help of their intestinal microorganisms. The first metagenomic analysis of the inhabitants of a termite gut provides insight into this feat of biomass-to-energy conversion.

The gut of wood-feeding termites is a tiny but astonishingly efficient bioreactor, in which microbes catalyse the conversion of lignified plant cell walls to fermentation products that drive the metabolism of their host. Molecular phylogenetic data have revealed the presence of hundreds of microbial species in this micro-litre-sized environment, but little is known about their functional diversity. On page 560 of this issue, Warnecke *et al.*¹ report an analysis of the metagenome of bacteria in the largest part of the gut — the hindgut paunch — of a termite species of the genus *Nasutitermes* (Fig. 1, overleaf).

Metagenomics is the burgeoning study of the entire genetic material from particular environments, and allows investigation of microorganisms that cannot be cultured in the laboratory. In this case, Warnecke *et al.* compiled an enormous amount of (unavoidably highly fragmented) sequence data from the total genomic DNA of the inhabitants of the termite hindgut paunch, and compared those data with known gene sequences in public databases. Such sequence comparisons

offer a wealth of information, but nevertheless can be deceptive. So, crucially, the authors also identified the major proteins secreted into the hindgut fluid, and went further by confirming the catalytic properties of selected gene products by expressing the genes in the bacterium *Escherichia coli*.

The main structural polysaccharides of wood (lignocellulose) are cellulose and various hemicelluloses: both are efficiently digested by termites. In lower (more primitive) termites, hydrolysis of cellulose and hemicelluloses is catalysed by flagellate protozoa housed in the hindgut paunch; bacteria are also present. Higher termites, which include *Nasutitermes* and most termite species worldwide, lack flagellates. Instead, their hindgut is host to a largely bacterial community, although the involvement of the bacteria in cellulose and hemicellulose digestion has not been clear².

The metagenomic analysis¹ of the *Nasutitermes* hindgut reveals a rich diversity of bacterial genes encoding hitherto unknown glycosyl hydrolases. These enzymes constitute more than 100 families of proteins that cleave

the glycosidic bond between carbohydrates, or between a carbohydrate and non-carbohydrate entity. The authors' sequence analysis suggests that many of the gene products fall within the glycosyl hydrolase families specializing in the degradation of cellulose and hemicelluloses. For several gene products, they also demonstrate cellulase activity *in vitro*. So far, bacteria with such cellulolytic power have not been isolated from the hindgut of higher termites; until recently, even their activity had remained undetected because the enzymes are not soluble but are associated with the particulate fraction of the gut content³.

Many of the genes identified by Warnecke *et al.*¹ could be assigned to one of two groups of bacteria — the fibrobacters and spirochetes — known to be abundant in the hindgut of *Nasutitermes* species. The fibrobacters from termite guts have not been cultured in the lab, but are known to be distant relatives of the fibre-degrading bacteria found in the intestines of cows and other ruminant animals⁴. The implication that members of the spirochetes are involved in cellulose digestion, however, was unexpected. This result is of similar importance to the earlier, equally surprising discovery that termite-gut spirochetes can carry out reductive acetogenesis⁵ (a mode of energy metabolism that results in the reduction of carbon dioxide to acetate). This finding answered the long-standing question regarding the identity of the organism responsible for one of the key metabolic activities in termite guts⁶.

Warnecke *et al.*¹ show that almost all of the



Figure 1 | Termite territory. This piece of 'carton nest' is typical of the structures built by *Nasutitermes* species. The nest material consists largely of the undigested residues of the termite diet of lignocellulose.

genes for an operational (acetyl-CoA) pathway for reductive acetogenesis seem to be present in the spirochetes of the *Nasutitermes* hindgut. This complements evidence for the predominance of spirochetes among the bacteria expressing this pathway in lower termite species. In *Nasutitermes*, however, the pathway seems to be peculiar, because the enzyme necessary for the production of formate is severely under-represented in the metagenome. It remains to be seen whether hydrogen — the central free intermediate during lignocellulose degradation in lower termites⁷ — is equally important in higher termites and, if so, whether the putatively spirochetal iron-only hydrogenases in the *Nasutitermes* hindgut are involved in hydrogen production or consumption. Other puzzling aspects of the metagenome are the presence of genes apparently encoding sensory hydrogenases, and an abundance of other genes putatively involved in signal transduction and chemotaxis in the hindgut community.

The gut microbiota of termites can theoretically convert a sheet of A4 paper into two litres of hydrogen. Figures such as this have focused attention on termite guts as a source of microorganisms and enzyme systems for the production of biofuels. The biotechnological challenges in this process are getting a grip on the initial step — the breakdown of the highly stable polymers of lignocellulose to microbial substrates — and redirecting the carbon and electron flow in the metabolic fermentations to useful products (such as ethanol or hydrogen). The largely unexplored biodiversity and biochemistry of the termite gut is a promising source of novel catalytic capacities. For example, it is intriguing that termites efficiently digest wood in the apparent absence of any classical lignin-degrading enzymes. The gut microbiota also holds the answers to puzzles

such as the unknown mechanisms that allow the release of reducing equivalents, formed during glycolysis, as free hydrogen against a high partial pressure.

The data set provided by Warnecke *et al.*¹ is a treasure trove for researchers with different interests. But it is just a beginning. For example, the presumably anaerobic microbiota of the hindgut paunch differs from that of the hindgut wall⁸ — a microhabitat characterized by the continuous influx of oxygen⁹. Microorganisms with different specializations are probably present in other gut regions, especially

in the extremely alkaline (pH 10–12) hindgut compartments characteristic of many higher termites². And functionally distinct microbial communities are to be expected in termites with other feeding habits, such as the soil-feeding species (for which nitrogen-rich soil peptides are important dietary components) or fungus-cultivating species (which thrive on material that has been partially delignified and is enriched in fungal biomass)².

Finally, another item for the research agenda is the diverse but specific association of bacteria and flagellate protozoa in the hindguts of lower termites¹⁰. These are not only a fruitful source of microbial biodiversity and functional novelties, but also hold the key to understanding the metabolic basis and evolutionary origin of a highly successful, tripartite symbiosis. ■

Andreas Brune is in the Department of Biogeochemistry, Max Planck Institute for Terrestrial Microbiology, 35043 Marburg, Germany.

e-mail: brune@mpi-marburg.mpg.de

1. Warnecke, F. *et al.* *Nature* **450**, 560–565 (2007).
2. Brune, A. in *The Prokaryotes Volume 1: Symbiotic Associations, Biotechnology, Applied Microbiology* 3rd edn (eds Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H. & Stackebrandt, E.) 439–474 (Springer, New York, 2006).
3. Tokuda, G. & Watanabe, H. *Biol. Lett.* **3**, 336–339 (2007).
4. Hongoh, Y. *et al.* *Appl. Environ. Microbiol.* **72**, 6780–6788 (2006).
5. Leadbetter, J. R., Schmidt, T. M., Graber, J. R. & Breznak, J. A. *Science* **283**, 686–689 (1999).
6. Brauman, A. *et al.* *Science* **257**, 1384–1387 (1992).
7. Pester, M. & Brune, A. *ISME J.* **1**, 551–565 (2007).
8. Nakajima, H., Hongoh, Y., Usami, R., Kudo, T. & Ohkuma, M. *FEMS Microbiol. Ecol.* **54**, 247–255 (2005).
9. Brune, A., Frenzel, P. & Cypionka, H. *FEMS Microbiol. Rev.* **24**, 691–710 (2000).
10. Brune, A. & Stiglitz, U. in *Molecular Basis of Symbiosis* (ed. Overmann, J.) 39–60 (Springer, Berlin, 2005).

ASTRONOMY

Sloan at five

Robert C. Kennicutt Jr

The Sloan Digital Sky Survey represents the most ambitious attempt yet to map out a slice of the sky. In the first five years of its existence, it has revealed cosmic structures on every conceivable scale.

Last month saw the publication of the paper¹ describing the fifth and final data release of the Sloan Digital Sky Survey (SDSS). The project has since been extended — a sixth data release was announced this summer — but this point marks the formal completion of the first phase of one of the most ambitious and productive astronomy projects ever undertaken.

Astronomical observatories generally achieve distinction by measuring the sky with unprecedented sensitivity, or by opening up a new region of the electromagnetic spectrum to measurements — hence the seemingly insatiable thirst of astronomers for ever-larger telescopes both on the ground and in space.

Sloan was different, using a single, relatively modest, 2.5-metre-aperture telescope located at the Apache Point Observatory, New Mexico. Its contribution, unparalleled so far in both the volume and the quality of the data produced, has been to map enormous swaths of sky, and thus harness the statistical power to be gained from precise measurements of millions of galaxies. In the first five years after its inception in 2000, it searched more than 8,000 square degrees of the northern sky — about a fifth of the entire sky — in five wavelength bands, bagging a booty of some 217 million objects, mostly galaxies, stars and asteroids (Fig. 1). Spectra were measured for around 675,000

galaxies, 215,000 stars and 90,000 quasars (galaxies hosting energetic outbursts in their centres), and all were placed in a freely available archive.

By any objective measure, Sloan has been a resounding scientific success. Two of the original science drivers for the project provide prime examples of discoveries both expected and unexpected. First, there was the desire to measure with unprecedented precision the web-like pattern of large-scale structure in the Universe. SDSS measurements have constrained the spectrum of the density fluctuations in the early Universe that led to this clustering², and provided a bedrock for the current 'concordance' cosmology of a Universe dominated by the mysterious substances known as dark energy and dark matter. Subsequent analysis of the data, together with those from the complementary Two Degree Field Galaxy Redshift Survey, also revealed even subtler features in the galaxy distribution that were produced by sound waves in the primordial Universe³. Measurements of these 'baryon acoustic oscillations' potentially offer a test of the nature of dark energy itself⁴.

Another original goal of the survey was the detection of quasars well beyond the limiting distances and look-back times (parametrized as 'redshift') then accessible. A total of 27 quasars above a redshift of 5.7, representing an epoch when the Universe was less than 10% of its present age, have been found. What came as an utter surprise was that ultraviolet light from these quasars rapidly dimmed with increasing redshift. This effect is caused by absorption through neutral intergalactic gas in those early times⁵, and marked the first observation of the transition from the cosmic 'dark ages', when the intergalactic medium was opaque to ultraviolet light, to the present cosmic epoch, in which the reionized intergalactic gas is transparent. The Sloan observations showed that the end of the dark ages occurred much later (and so at lower redshifts) than had been expected, and have fuelled theoretical attempts to understand the nature and origins of the objects that caused the reionization.

The catalogue of SDSS discoveries spans the gamut of scales, from the structure of the Universe in bulk to the faintest stars and to objects in our Solar System. Wide-area imaging has revealed orbiting streams of stars in the halo of our own Milky Way galaxy that are the debris of accreted satellite galaxies — direct evidence in Earth's immediate neighbourhood for the hierarchical merging of structures that is a ubiquitous process in the concordance 'cold dark matter' cosmology^{6,7}. Not much farther away is a parade of newly discovered dwarf galaxies — most of them far less massive than any previously known, and some in the process of being accreted and shredded apart by the Milky Way⁸.

At the other extreme, Sloan has also discovered a few stars that seem to be escaping from



Figure 1 | Stars in a whirl. The Whirlpool galaxy M51a and its companion, the dwarf galaxy NGC 5195, two of the more than 200 million objects imaged by the Sloan Digital Sky Survey. Some 30 million light years away in the constellation Canes Venatici, M51a is around half the size of our own Galaxy, with a radius of about 38,000 light years and a mass of 160 billion Suns. Its distinctive spiral structure is thought to have been caused by the slow passage of NGC 5195 through its main disk over hundreds of millions of years.

the Milky Way, as a result of violent gravitational encounters in its central regions⁹. From the faint red objects discarded as quasars at high redshifts, a new class of faint stars and substellar objects, including the previously discovered 'brown-dwarf' objects intermediate between the largest gas-giant planets and the smallest viable stars, has arisen¹⁰. And, really in our own back yard, analysis of multi-epoch SDSS images has produced a large sample of asteroids, together with information on their sizes and reflection properties, allowing them to be grouped into orbital families with a common origin¹¹.

The story is far from finished. The SDSS collaborators are currently completing a second-generation survey, which extends the sky coverage of the original survey, maps the northern plane of the Milky Way, and includes a multiple-epoch survey for supernovae and other time-variable phenomena. Plans are in place for a third-generation survey that will include the more detailed measurement of baryonic acoustic oscillations, a further extension of the imaging of the Galactic plane, a spectroscopic survey of the Galactic halo, and a search for extrasolar planets.

In the meantime, the original survey data remain a vast archive that in many aspects has hardly been touched. Trillions of pixels of imaging data are only beginning to be

analysed in depth, and mating the wide sweep of the Sloan observations to other new surveys at ultraviolet, infrared, X-ray and radio wavelengths will bring immense new diagnostic power to bear.

By altering the traditional interactions between a telescope, its data and communities of astronomers, Sloan has transformed astronomy. The fact that its voluminous data archive is available to any aspiring astronomer who is skilled and conscientious enough to extract information from it — roughly half of the 1,500 refereed papers based on SDSS data up to now have come from authors outside the project itself, with that proportion rising — is indeed a legacy to be celebrated. ■

Robert C. Kennicutt Jr is at the Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK. e-mail: robk@ast.cam.ac.uk

- Adelman-McCarthy, J. K. *et al. Astrophys. J. Suppl. Ser.* **172**, 634–644 (2007).
- Tegmark, M. *et al. Astrophys. J.* **606**, 702–740 (2004).
- Eisenstein, D. J. *et al. Astrophys. J.* **633**, 560–574 (2005).
- Eisenstein, D. J. *et al. Astrophys. J.* **664**, 675–679 (2007).
- Fan, X. *et al. Astron. J.* **123**, 1247–1257 (2002).
- Newberg, H. J. *et al. Astrophys. J.* **561**, 245–274 (2002).
- Belokurov, V. *et al. Astrophys. J.* **642**, L137–L140 (2006).
- Belokurov, V. *et al. Astrophys. J.* **654**, 897–906 (2007).
- Brown, W. R. *et al. Astrophys. J.* **622**, L33–L36 (2005).
- Hawley, S. L. *et al. Astron. J.* **123**, 3409–3427 (2002).
- Ivezic, Z. *et al. Astron. J.* **124**, 2943–2948 (2002).

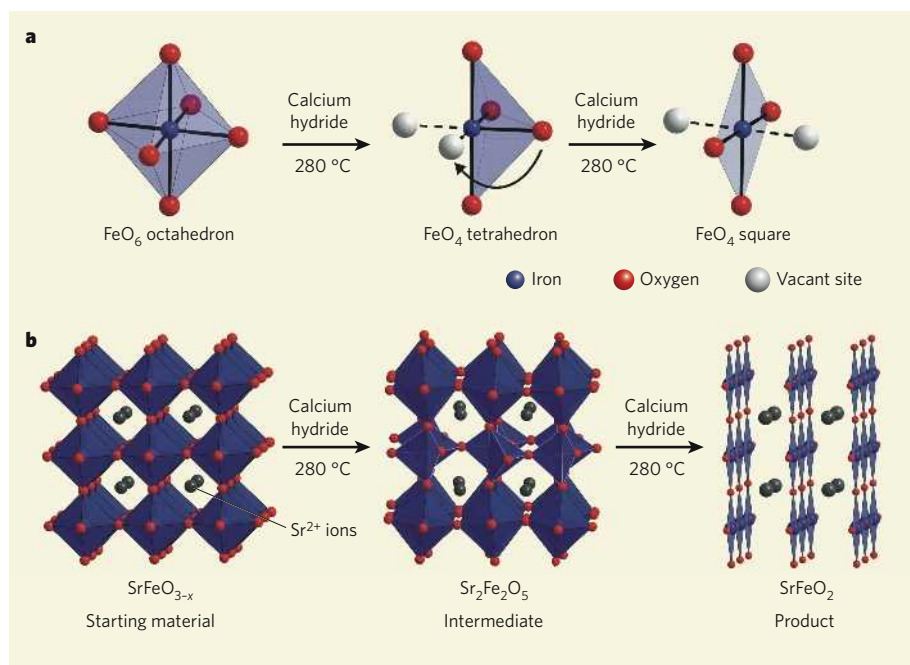


Figure 1 | Ion movement in a complex metal oxide. Tsujimoto *et al.*¹ report the preparation of SrFeO₂ (Sr is strontium, Fe is iron), a complex metal oxide with an unusual arrangement of ions. **a**, The starting material (SrFeO_{3-x}, where *x* is about 0.125) contains FeO₆ octahedra. On heating with calcium hydride, some of the oxygen ions (known as oxide ions) are lost, so that intermediate FeO₄ tetrahedra form. One of the remaining oxide ions then moves to a vacant site left behind by an ion (indicated by the arrow), forming an FeO₄ square. **b**, The crystal lattice of the starting material is an array of FeO₆ octahedra, in which each oxygen is shared between two iron atoms. Strontium ions (Sr²⁺) fit in between the rows of octahedra. In the first step of the reaction, loss of some of the oxide ions leads to an intermediate compound, Sr₂Fe₂O₅, consisting of alternating rows of FeO₆ octahedra and FeO₄ tetrahedra. In the second step, more oxide ions are lost and some of the remaining oxide ions change position. The SrFeO₂ product thus forms as sheets of FeO₄ squares, interleaved with strontium ions. The iron and strontium ions retain their positions throughout the process.

consider the differences between solid-state reactions and those that occur in solution. Reactions in solution can be performed at low temperatures, because molecular diffusion occurs easily and the reacting molecules don't require much energy to mix together. Under these conditions, if a molecule can take part in several reactions, only the one with the lowest energy barrier to activation tends to occur. The product of such a 'kinetically controlled' reaction is the one that forms fastest, and is not necessarily the one that is most stable. It is therefore possible to control reactions in solution so that they occur only at specific parts of a molecule. By performing stepwise transformations on individual chemical groups, a product can be prepared that has a controlled composition and structure that are clearly related to those of the starting compound.

But nearly all complex metal oxides are prepared at high temperatures (typically greater than 1,000 °C). This is because no solvent is used to aid diffusion, yet the reacting ions must travel large distances (of the order of micrometres) to form the products. At these temperatures, enough energy is available to allow the occurrence of reactions that have high energy barriers. Given a choice of reaction pathways, the most favourable one is that which yields the most thermodynamically stable configuration of atoms. In such

thermodynamically controlled systems, the product generally does not conserve any of the structural features of the reactants (unlike reactions in solution), and it is therefore much more difficult to direct the course of the reaction⁵.

Tsujimoto *et al.*¹ overcome these limitations in their preparation of SrFeO₂. Their starting material is a complex metal oxide that contains Fe⁴⁺ ions (SrFeO_{3-x}, where *x* is about 0.125). The authors form their unusual product by removing an oxide ion from the starting material, a process that is coupled to a redox reaction in which Fe⁴⁺ ions are converted into Fe²⁺ ions.

The overall process uses a recently discovered reagent (calcium hydride) for the kinetically controlled removal of oxygen from oxides⁶, and it occurs at the remarkably low temperature of 280 °C. These conditions provide insufficient thermal energy to rearrange the structure of the starting material completely — only the relatively mobile oxide ions can change position (Fig. 1). The strontium and iron ions in the product retain the positions they held in the starting material. The most thermodynamically stable products — iron metal and strontium(II) oxide — do not form, because the required long-range diffusion for the process is too slow at this temperature. The less stable SrFeO₂ forms instead, because this is a faster reaction.

Nevertheless, the reaction pathway that

leads to SrFeO₂ (Fig. 1) is unexpected. An intermediate (Sr₂Fe₂O₅) is formed first, as oxide ions are removed from the starting material. This intermediate consists of alternating sheets of FeO₄ tetrahedra and FeO₆ octahedra. Conversion of the tetrahedra into the square planes of the final product requires that the oxide sites vacated in the formation of the intermediate be refilled with other oxide ligands. This is a crucial observation, because it demonstrates that all the oxide ions are mobile, not just those being removed from the system.

The discovery that oxide ions can be mobile at relatively low temperatures, albeit in the presence of a strong chemical driving force, opens up a host of synthetic possibilities — for example, the strong magnetic interactions seen in SrFeO₂ could be modified in a controlled way by making complex oxides of different transition metals. But the practical applications are just as exciting — high oxide-ion mobility is required for several emerging technologies, most notably solid-oxide fuel cells⁷. So although Tsujimoto and colleagues' discovery¹ may occur only at an atomic level, its ramifications could extend far more widely.

Michael A. Hayward is in the Inorganic Chemistry Laboratory, Department of Chemistry, University of Oxford, South Parks Road, Oxford OX1 3QR, UK. Matthew J. Rosseinsky is in the Department of Chemistry, University of Liverpool, Liverpool L69 7ZD, UK.

e-mails: michael.hayward@chem.ox.ac.uk; m.j.rosseinsky@liverpool.ac.uk

1. Tsujimoto, Y. *et al.* *Nature* **450**, 1062–1065 (2007).
2. Blundell, S. *Magnetism in Condensed Matter* (Oxford Univ. Press, 2001).
3. Rao, C. N. R. & Raveau, B. *Transition Metal Oxides: Structure, Properties, and Synthesis of Ceramic Oxides* 2nd edn (Wiley-VCH, Weinheim, 1998).
4. Cava, R. J. *J. Am. Ceram. Soc.* **83**, 5–28 (2000).
5. Stein, A., Keller, S. W. & Mallouk, T. E. *Science* **259**, 1558–1564 (1993).
6. Hayward, M. A. *et al.* *Science* **295**, 1882–1884 (2002).
7. Atkinson, A. *et al.* *Nature Mater.* **3**, 17–27 (2004).

Corrections

■ The News & Views article "Venus: Express dispatches" by Andrew P. Ingersoll (*Nature* **450**, 617–618; 2007) contained the erroneous statement that Venus's equator is warmer than the poles at altitudes above 65 km. It is colder.

■ There was an incorrect reference citation in the article "Microscopy: Elementary resolution" by Christian Colliex (*Nature* **450**, 622–623; 2007). In the statement "The first experimental maps are now demonstrating the importance of refining descriptions of electron-matter interactions²", the correct citation is not reference 2 but reference 12 (M. Bosman *et al.* *Phys. Rev. Lett.* **99**, 086102; 2007).

■ In the article "Astronomy: Sloan at five" by Robert C. Kennicutt Jr (*Nature* **450**, 488–489; 2007), we should clarify that the Sloan Digital Sky Survey was used only to select candidate stars for the spectroscopic observations that led to the discovery cited in reference 9 (W. R. Brown *et al.* *Astrophys. J.* **622**, L33–L36; 2005).

EARTH SCIENCE

Sediment en route to oblivion

Philip A. Allen

Sudden collapses of the sea floor can generate oceanic sediment flows that dwarf the global annual sediment input from rivers. Such flows can travel great distances, and undergo transformation along the way.

The seabed of the oceans is prone to failure, generating high-density sediment flows of immense volume that can run for hundreds to thousands of kilometres from the continental margin to the deep oceanic plains. On page 541 of this issue, Talling and co-workers¹ report on one such flow, which occurred about 60,000 years ago off northwest Africa.

The authors estimate that, in a period of hours or at most days, 22.5×10^{13} kilograms of sediment was transported into the ocean depths — ten times more than is annually delivered into the oceans by rivers. The sediment flow started as a submarine landslide; it moved downslope as a turbulent suspension as it was funnelled through the narrows of the Agadir canyon (Fig. 1); it was then transformed into a cohesive mass of debris by a slight reduction in the slope of the seabed. The work is significant in documenting the way in which seabed-hugging sediment flows undergo transformations in their flow behaviour. It also highlights the hazard that these huge flows pose to seabed installations, and to coastal communities given the potential of such flows to trigger a tsunami.

It is well over a hundred years since the Swiss limnologist François-Alphonse Forel (1841–1912) detected sediment flows in the Lake of Geneva caused by floods of the River Rhône. Later, in 1929, an earthquake triggered a sediment flow on the Grand Banks off Newfoundland that broke successive telecommunication cables on the seabed, and so demonstrated the existence of fast-moving, long-distance sediment flows across the ocean floor. Today, improved imaging techniques, combined with recovery of cores of sediment that sample hundreds of thousands of years of seabed activity², are transforming our view of the deep ocean floor. The picture is a very dynamic one.

The surface processes on continental margins such as that of northwest Africa are increasingly well understood, and show many signs of seabed instability, from the rock avalanches derived from collapse of the volcanic islands of the Canaries³, to the extensive swaths of gigantic debris units, such as the Saharan flow⁴, to the fretting of the continental margin by channels that focus dilute turbulent suspensions on their way to the depths of the Madeira abyssal plain⁵ (Fig. 1).

Talling and co-workers¹ identified a deposit, or bed, of sediment and debris that is less than 2 metres thick but that can be followed for a staggering distance of more than 1,000 kilometres. Over part (250 kilometres) of this distance, the bed contains a cohesive unit of debris flow (debrite) encased in sand (turbidite), attributed to the activity of more dilute turbulent suspensions known as turbidity currents. The onset of debrite deposition is correlated with a small but significant reduction in seabed slope from 0.05° to 0.01° . In other words, the debrite and the encasing turbidite are co-genetic, or 'linked'⁶, appearing like a sandwich⁷, and providing evidence of a downslope transformation of a turbidity current into a cohesive debris flow triggered by a reduction of slope. Conventional thinking, by contrast, is that debris flows most commonly evolve from the disintegration of moving landslide material, and in turn became more dilute and turbulent during downslope transport as they take in water⁸.

The processes recognized on the northwestern margin of Africa are assumed to be typical of the main way in which mass is transferred to the deep ocean. In the past, sedimentologists recognized trends in the vertical superimposition of beds of turbidites and debris flows, but the lateral persistence and variability of beds were largely unknown. One way of bridging this gap in understanding the three-dimensional geometry of deep marine deposits is to build a statistical model using the scaling relationships between bed thickness, lateral extent and volume^{9–11}. A lateral transformation of debrite to co-genetic turbidite would be troublesome to these scaling relationships.

Another way to make progress is to document patterns in the architecture of deep marine deposits using selected examples. The Marnoso arenacea formation — once such a deposit but now evident on land as part of the Apennine mountains in northern Italy — is a celebrated case¹². Such documentations offer the possibility of assessing the 'run-out' distances of individual flows. In the case of co-genetic flows, it may be possible to answer the intriguing question of whether gigantic run-out distances are aided by flow-transformation processes⁶.

Deep marine deposits commonly contain valuable resources such as hydrocarbons, so energy companies have been keen to map them in three dimensions. Three-dimensional heterogeneity affects the ability to extract liquid and gaseous resources from these deposits, and it primarily originates from the interbedding of two types of deposit, sandstones and mudstones, left by the main body and dilute tail of turbidity currents respectively. The lateral transformation from a cohesive debrite (which has low porosity) to a turbidite (with high porosity) is another source of heterogeneity. If such a transformation can be triggered by a slight reduction in slope, as Talling and co-workers¹ suggest, the question arises of whether it is possible to predict such

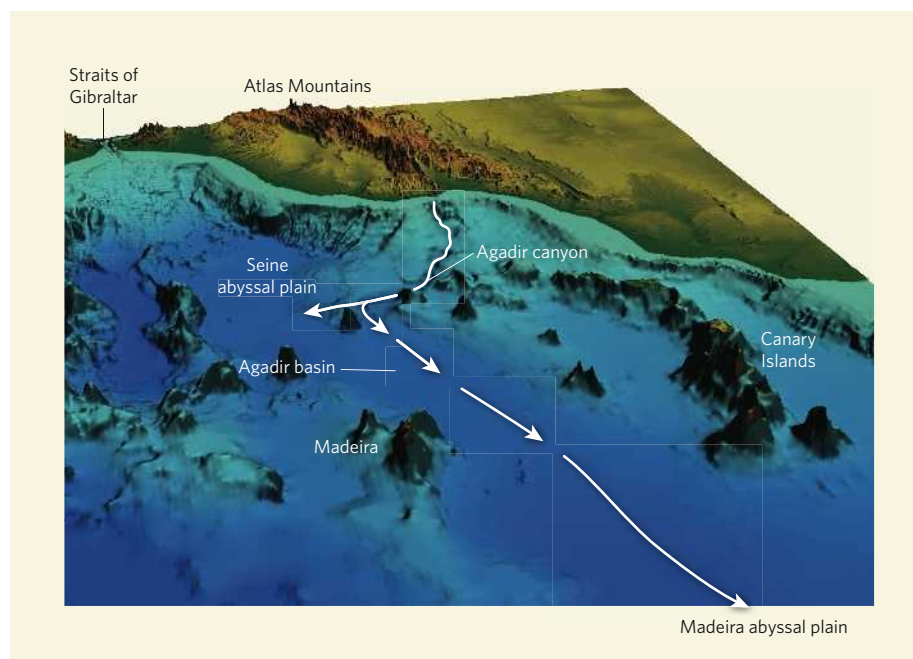


Figure 1 | Ocean bathymetry off the northwest coast of Africa. The pathway of the sediment flow responsible for the deposit (called bed 5) investigated by Talling *et al.*¹ is depicted by arrows. The flow started as a submarine landslide, and passed through the Agadir canyon. Part of it was deposited on the Seine abyssal plain; part travelled through the Agadir basin en route for the Madeira abyssal plain. These plains lie more than 4 kilometres beneath the ocean surface, and the Madeira abyssal plain is more than 1,500 kilometres distant from the site of the initial landslide. (Bathymetric image courtesy of T. Le Bas.)

slope-induced changes and their effects on flow behaviour. Given that many deep marine deposits occur in sedimentary basins affected by faulting, however, and that high-density sediment flows commonly enter semi-confined basins with lateral slopes, the answer in most cases is probably 'no'.

The bed-by-bed correlation carried out by sedimentologists on surface exposures such as the Marnoso arenacea, or on cores from below the ocean floor as documented by Talling *et al.* in their Supplementary Information¹, is a painstaking business. The reward for such endeavours is an appreciation of the true

variability of deep marine sediments and of the mechanics of their deposition. The large volumes of sediment en route to oblivion in the dark ocean basins reveal a complexity of seabed processes unimagined even by the visionary Forel, who likened his beloved Swiss lakes to small oceans.

Philip A. Allen is in the Department of Earth Science & Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.
e-mail: philip.allen@imperial.ac.uk

1. Talling, P. J. *et al.* *Nature* **450**, 541–544 (2007).

2. Weaver, P. P. E., Rothwell, R. G., Ebbing, J., Gunn, D. E. & Hunter, P. M. *Mar. Geol.* **109**, 1–20 (1992).
3. Masson, D. G. *et al.* *Earth Sci. Rev.* **57**, 1–35 (2002).
4. Gee, M. J. R., Masson, D. G., Watts, A. B. & Allen, P. A. *Sedimentology* **46**, 317–335 (1999).
5. Masson, D. G. *Basin Res.* **6**, 17–33 (1994).
6. Haughton, P. D. W., Barker, S. P. & McCaffrey, W. D. *Sedimentology* **50**, 459–482 (2003).
7. McCaffrey, W. & Kneller, B. *Bull. Am. Assoc. Petrol. Geol.* **85**, 971–988 (2001).
8. Mohrig, D., Whipple, K. X., Hondzo, M., Ellis, C. & Parker, G. *Geol. Soc. Am. Bull.* **110**, 387–394 (1998).
9. Rothman, D. H., Grotzinger, J. P. & Flemings, P. J. *Sediment. Res.* **64**, 59–67 (1994).
10. Malinverno, A. *Basin Res.* **9**, 263–274 (1997).
11. Sinclair, H. D. & Cowie, P. A. J. *Geol.* **11**, 277–299 (2003).
12. Amy, L. A. & Talling, P. J. *Sedimentology* **53**, 161–212 (2007).

CARBON CYCLE

Marine manipulations

Kevin R. Arrigo

The effect of increasing levels of atmospheric carbon dioxide on carbon uptake in and export from the upper ocean is one of the big questions in environmental science. But it can be tackled experimentally.

Marine phytoplankton are major players in the carbon cycle, accounting for about 50% of the global biological uptake of carbon dioxide¹. Near the ocean surface, these single-celled organisms use light energy to convert CO₂ into organic molecules for building cellular structures and driving their metabolism. Some of this organic carbon eventually sinks into the deep ocean, where most of it is either converted back to CO₂ or sequestered in sediments. This 'biological pump' effectively removes CO₂, a greenhouse gas, from the atmosphere for hundreds to millions of years.

On page 545 of this issue, Riebesell *et al.*² describe evidence that the biological pump may become stronger at elevated concentrations of CO₂ in the atmosphere, and thus provide a negative feedback on increasing atmospheric CO₂. According to their calculations, that feedback has accounted for about 10% of the extra CO₂ pumped into the atmosphere since pre-industrial times (the past 200 years or so).

Since industrialization, atmospheric CO₂ has risen from about 280 parts per million (p.p.m.) to more than 385 p.p.m., increasing by some 2 p.p.m. per year during the past decade. Each year, approximately 25–30% of anthropogenic CO₂ enters the surface ocean, where it increases both the concentration of dissolved inorganic carbon (DIC) and acidity. The latter has potentially adverse consequences for phytoplankton that require calcium carbonate to build their shells. Although the oceans are Earth's largest reservoir for DIC, only about 1% is in the form of CO₂, the molecule required by the photosynthetic enzyme rubisco. At the low CO₂ concentrations typical of sea water, rubisco operates at rather low efficiency³. So increasing ambient concentrations of CO₂ in sea water could boost

photosynthetic efficiency and increase biological uptake of anthropogenic CO₂, just as some marine phytoplankton use intracellular carbon-concentrating mechanisms to increase their photosynthetic capacity.

This is the context for Riebesell and colleagues' research² into how phytoplankton might respond to increasing CO₂ concentrations. They conducted CO₂ manipulations in large cylindrical enclosures called mesocosms

that were placed in a fjord in southern Norway and extended from the surface to a depth of approximately 9–10 metres. Although this approach is complex and logistically difficult, the advantage is that mesocosms are exposed to the same environmental influences as the surrounding waters, making them reasonable analogues for natural systems. And they can be manipulated experimentally. In Riebesell and colleagues' study, phytoplankton were grown in different mesocosms with the partial pressure of CO₂ adjusted to simulate the present (350 μ atm) or projected future (2 \times present CO₂, 700 μ atm, and 3 \times present CO₂, 1,050 μ atm) atmosphere.

What the authors found was intriguing. Uptake of CO₂ by phytoplankton (mainly bloom-forming diatoms and coccolithophores) in the 2 \times CO₂ and 3 \times CO₂ treatments was 27% and 39% higher, respectively, than in the present-day CO₂ treatment. But the additional

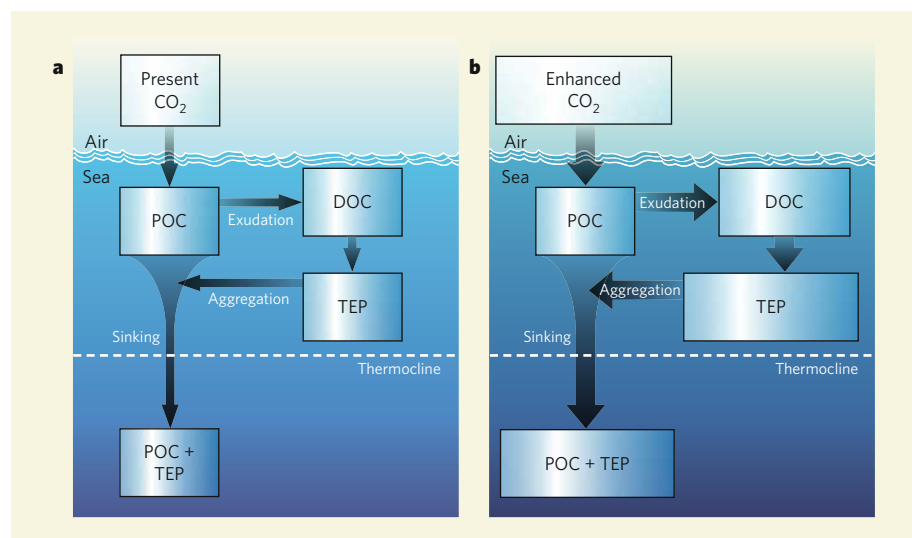


Figure 1 | Carbon dioxide in the atmosphere, and organic carbon in the ocean. **a**, The size, relative to part **b**, of the different pools and fluxes of organic carbon under present levels of atmospheric CO₂ (POC, particulate organic carbon; DOC, dissolved organic carbon; TEP, transparent exopolymeric particles). The thermocline is an abrupt temperature discontinuity that acts as a barrier between the upper mixed ocean and deeper waters. **b**, According to the results of Riebesell *et al.*², uptake of CO₂ by phytoplankton increases at enhanced CO₂ concentrations (thicker arrow). Exudation of DOC from the pool of POC (primarily phytoplankton) also increases, although the POC pool itself remains unchanged. This extra DOC coalesces to form a larger pool of TEP that facilitates increased POC aggregation and enhances sinking fluxes. Thus, the flux of carbon from the atmosphere to the deep ocean is increased at higher atmospheric concentrations of CO₂.

CO₂ removed from surface waters at elevated CO₂ was not balanced by increases in particulate organic carbon (POC) in the surface layer. Furthermore, the loss of nitrate from the surface waters was the same in all three CO₂ treatments, indicating that the ratio of carbon to nitrogen uptake increased at higher CO₂ concentrations whereas the cellular carbon/nitrogen ratio of the phytoplankton remained unchanged.

This result suggests that, although higher ambient CO₂ concentrations increased CO₂ uptake by phytoplankton, the additional carbon incorporated into cells was rapidly lost as dissolved organic carbon (DOC). However, although DOC concentrations in the mesocosms increased, these were insufficient to balance the measured CO₂ deficits. In nature, organic molecules excreted from phytoplankton (for example, as DOC), or otherwise lost as these organisms die or are grazed, can coalesce to form semi-solid structures called transparent exopolymer particles (TEPs). These structures are sticky and facilitate the aggregation and increased sinking speeds of other particulate matter. In the mesocosms, TEP concentrations increased fourfold during the experiment (the carbon content of these TEPs is not presented by Riebesell *et al.*).

The authors propose that accumulations of TEPs in the elevated-CO₂ treatments facilitated aggregate formation, increasing the flux of particulate matter from the mesocosm surface. Coupled with higher DOC production, this may explain why POC did not increase in the elevated CO₂ treatments. Thus, it seems that increased CO₂ uptake fuelled by higher CO₂ concentrations was rapidly converted to DOC and TEPs, and any additional carbon incorporated into POC was lost from the surface of the mesocosm owing to increased particle aggregation and sinking (Fig. 1). Assuming that their results are representative of the larger ocean, increased atmospheric CO₂ may lead indirectly to increased particle fluxes from the surface ocean to depth, providing a negative feedback to increasing atmospheric CO₂ concentrations. Unfortunately, the authors did not measure POC sinking fluxes in their mesocosms to confirm this link.

Nevertheless, there are some notable conclusions to be drawn from this study. First, although CO₂ uptake by phytoplankton may be stimulated in a high-CO₂ world, this negative feedback will only partly offset expected increases in atmospheric CO₂. In fact, Riebesell *et al.* perform some clever calculations to show that the CO₂-enhancement effect they identified has probably reduced the rise in atmospheric CO₂ by only 11 μatm (about 10%) since the dawn of the industrial revolution.

More importantly, their study provides a vivid example of the fact that ocean biology is not in steady-state and that fundamental biological and biogeochemical processes are likely to respond to climate change, resulting in either positive or negative feedbacks that

are difficult to predict. One positive feedback between biology and climate has already been identified, whereby future increases in stratification of the Southern Ocean could favour types of phytoplankton that have a reduced capacity to take up CO₂ (ref. 4). Conversely, increased CO₂ has been shown to enhance fixation of free nitrogen, thereby relaxing nutrient limitation by nitrogen availability and increasing CO₂ uptake⁵. Riebesell and colleagues document another negative feedback whereby CO₂ use by the dominant bloom-forming groups of phytoplankton could increase as atmospheric levels of CO₂ rise. Neither these, nor other possible non-steady-state biological feedbacks, are currently

accounted for in models of global climate — a potentially serious omission, given that the biological pump is responsible for much of the vertical CO₂ gradient in the ocean. ■

Kevin R. Arrigo is in the Department of Geophysics, Stanford University, Stanford, California 94305-2215, USA.

e-mail: arrigo@stanford.edu

1. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. *Science* **281**, 237–240 (1998).
2. Riebesell, U. *et al.* *Nature* **450**, 545–548 (2007).
3. Raven, J. A. & Johnston, A. M. *Limnol. Oceanogr.* **36**, 1701–1714 (1991).
4. Arrigo, K. R. *et al.* *Science* **283**, 365–367 (1999).
5. Barcelos e Ramos, J., Biswas, H., Schulz, K. G., LaRoche, J. & Riebesell, U. *Global Biogeochem. Cycles* **21**, GB2028 (2007).

HIGH-TEMPERATURE SUPERCONDUCTIVITY

Schizophrenic electrons

Christian Pfleiderer and Rudi Hackl

The split personality of the conduction electrons in one high-temperature superconductor might indicate that periodic modulations of their spin and charge density are a general feature of these mystifying materials.

In simple metals, conduction electrons undergo well-understood phase transitions: they can become superconducting or ferromagnetic, or acquire periodic modulations of their spin and charge density. But just over 20 years ago, high-temperature superconductors were discovered, a class of materials in which the conduction electrons behave almost entirely outside these traditional models of order. Or do they? On page 533 of this issue, LeBoeuf *et al.*¹ report measurements of a classic high-temperature superconductor, yttrium barium copper oxide (YBCO), that hint at the presence of order in the form of a periodic density wave of conduction electrons. Together with similar sightings in other materials^{2,3}, this finding might indicate that this form of order is shared by all high-temperature superconductors.

High-temperature superconductivity — the conduction of electric current without resistance at temperatures of up to halfway between absolute zero and room temperature — presents a tremendous challenge to our understanding. All known high-temperature superconductors are copper-oxide (cuprate) materials. The highest superconducting transition temperature (the temperature above which superconducting behaviour is lost) occurs in a regime where the concentration of charge carriers in the material is somewhere between that of its magnetically ordered insulator state and that of its non-magnetic metallic state.

The charge-carrier concentration in YBCO and many other cuprates is controlled by 'doping' through the addition or subtraction

of a small number of oxygen atoms. One might expect not just the number but also the placing of these atoms (and the crystal structure in general) to influence the transition temperature. But oddly, the superconductivity seems fairly insensitive to the precise crystallographic arrangement. What is worse, the energy scales of all other phenomena that might cause the superconducting behaviour — the frequency of vibrations in the material's crystal structure, the speed of movement of conduction electrons, the rate at which islands of magnetization and modulations of charge density form and decay in the material — are similar, making it difficult to single out any particular one as the culprit. Any reasonable suggestion of primary forms of electronic order underlying the superconductivity would therefore be gratefully received.

In pursuit of such order, LeBoeuf *et al.*¹ set out to study the nature of charge carriers in YBCO superconductors. In solids such as YBCO, electrons reside in energy bands that form when the atomic orbitals of participating atoms overlap, establishing a relationship between an electron's energy and its momentum. The highest energy up to which bands are filled determines a surface, the Fermi surface, in a three-dimensional momentum space. As long as a band is occupied by just a few electrons, these behave as if they are essentially free, and can contribute to a flow of electric current. As soon as a band is completely occupied, this contribution ceases. But a strange thing happens if just a few electrons are missing from an otherwise fully occupied band. In this case, it is as if the missing electrons are free to

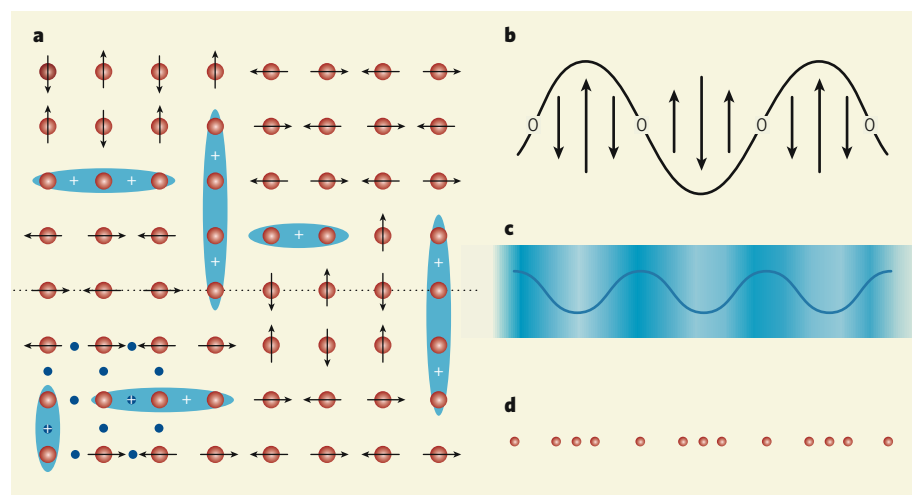


Figure 1 | Density wave in YBCO. **a**, In the copper–oxygen plane of the high-temperature superconductor yttrium barium copper oxide (YBCO; copper atoms are shown in red; for clarity, oxygen atoms (blue) are shown only in the lower left corner), positively charged holes introduced by doping (+) organize in stripe-like regions that break up long-range magnetic order. (In YBCO's insulating state at low doping, spins are arranged with antiferromagnetic order, with adjacent spins pointing in opposite directions.) Under doping and at higher temperatures, this pattern is not rigid, but fluctuates ever more rapidly. **b**, **c**, A snapshot of the modulation of the spin (**b**) and charge (**c**) densities along a cut indicated by the dashed line in (**a**). The charge modulation has twice the periodicity of the spin modulation. **d**, If the charge density modulation varies slowly enough, the copper and oxygen atoms may follow, exhibiting a new periodicity. A possible explanation for LeBoeuf and colleagues' Hall-effect findings¹ is that these processes can show up as a reconstruction of the Fermi surface: if the electrons can complete a full circle in the Hall field before the stripe pattern changes, the quantum oscillations that the authors observe will arise.

move through the crystal, producing a flow of positive electric charges, or 'holes'.

The authors¹ investigate a point in YBCO known as 1/8 filling, which lies slightly below the doping level at which the superconducting transition temperature is highest. In this configuration, there is, on average, one additional hole for every eight copper atoms in the copper–oxygen plane of the material. Together with the underlying magnetism of the copper atoms, this leads to a periodic, stripe-like modulation of the otherwise homogeneous charge distribution. Several anomalous characteristics associated with incipient spin or charge order had already been found around 1/8 filling in practically all cuprates^{2–4}.

The new experiments¹ were in several respects extreme, requiring both great advances in materials preparation techniques^{5,6} and the maintenance of temperatures near absolute zero at magnetic fields almost one million times stronger than Earth's magnetic field. The chosen probe was the Hall effect, a phenomenon first noted by the American physicist Edwin Hall in 1879. This effect occurs when electric currents are deflected in a weak magnetic field by means of the Lorentz force. Electrons and holes move in opposite directions, causing a build-up of charge — and therefore voltage — at right angles to both the direction of current flow and the magnetic field. The sign of the Hall voltage shows whether the current is predominantly carried by electrons (negative voltage) or by holes (positive).

In strong magnetic fields and in samples sufficiently clear of crystal defects, the electrons

and holes can race around in complete circles. The Russian physicist Lev Landau pointed out in 1930 that the energy associated with this circular motion is quantized. The Hall voltage then undergoes so-called quantum oscillations when the energy of the quantized race-tracks matches the energy corresponding to an extreme cross-sectional area of the material's Fermi surface. Extremely pure samples are needed to observe this subtle, but important, facet of the electronic structure, because the quantum oscillations are extremely sensitive to defects that kick the electrons out of their race-track.

LeBoeuf and colleagues' new offering¹ closely follows their previous work⁷, in which they also used the Hall technique to probe YBCO superconductors. So what's new? Before, they observed quantum oscillations in an overall negative Hall voltage, but attributed them to hole pockets. The sign of the voltage must either have been intrinsic to the hole-doped material (in which a positive Hall voltage would be expected), but of unidentified origin, or have indicated the presence of vortices of supercurrents circulating around a flux line of the applied magnetic field⁸. Only now have the authors pinned down the change of sign of the Hall signal from positive to negative as being intrinsic. They have thus identified an inherent contradiction: the charge carriers in YBCO have a split personality between electron-like and hole-like behaviours. This observation immediately raises the questions of where the electron pockets come from, and why there are no quantum oscillations of the holes.

One obvious answer to this question is provided by the order that occurs in the spin and charge density of conduction electrons in the metal chromium⁹. Here, electron pockets may be produced by the 'reconstruction' of the Fermi surface through a density wave. Figure 1 shows what, by analogy, might be going on in YBCO: the periodicity initially imposed by the charge modulation at 1/8 filling could eventually introduce a new periodicity in density, and thus in the lattice, and hence a reconstructed Fermi surface. An abundance of experiments suggests that such a density wave would be dynamic, yet, for the Hall effect and quantum oscillations, the electron-like behaviour may still be coherent, whereas the hole-like behaviour (perhaps because of the underlying timescales) is incoherent, and thus invisible. Even bearing in mind that the conjecture¹ of a density wave is based on data taken at very strong magnetic fields, the prospect that a density wave can connect a modulation of charge or spin density in real space with a modulation seen in the Fermi surface, a phenomenon of momentum space, is truly exciting.

This work also offers an opportunity for comparison with other unconventional, but low-temperature, superconductors. Perhaps the fastest growing class are the f-electron or heavy-fermion superconductors, of which more than 30 have so far been identified. In many of those systems in which quantum oscillations have been observed, a reconstruction of parts of the Fermi surface topology has been spotted precisely when applied pressure tunes the superconductivity to be strongest¹⁰. Just as LeBoeuf *et al.*¹ might well have shown for YBCO, the conduction electrons in f-electron superconductors seem to have a split personality — not between electron-like and hole-like behaviours, but between itinerant and localized electron states. Besides their impact on high-temperature superconductivity *per se*, the latest results also hit on a more general issue: why does superconductivity seemingly emerge in the presence of schizophrenic conduction electrons?

Christian Pfleiderer is at the Lehrstuhl für Experimentalphysik E21, Technische Universität München, D-85748 Garching, Germany. Rudi Hackl is at the Walther-Meißner-Institute of the Bayerische Akademie der Wissenschaften, D-85748 Garching, Germany.
e-mails: christian.pfleiderer@frm2.tum.de; hackl@wmi.badw.de

1. LeBoeuf, D. *et al.* *Nature* **450**, 533–536 (2007).
2. Tranquada, J. M. preprint available at www.arxiv.org/abs/cond-mat/0512115 (2005).
3. Kivelson, S. A. *et al.* *Rev. Mod. Phys.* **75**, 1201–1241 (2003).
4. Castellani, C., Di Castro, C. & Grilli, M. *Z. Phys. B* **103**, 137 (1997).
5. Erb, A., Walker, E. & Flükiger, R. *Physica C* **258**, 9–20 (1996).
6. Liang, R. *et al.* *Physica C* **336**, 57–62 (2000).
7. Doiron-Leyraud, N. *et al.* *Nature* **447**, 565–568 (2007).
8. Julian, S. R. & Norman, M. R. *Nature* **447**, 537–539 (2007).
9. Fawcett, E. *Rev. Mod. Phys.* **60**, 209–283 (1988).
10. Settai, R., Takeuchi, T. & Onuki, Y. *J. Phys. Soc. Jpn* **76**, 051003 (2007).

FEATURE

Grand challenges in chronic non-communicable diseases

The top 20 policy and research priorities for conditions such as diabetes, stroke and heart disease.

Abdallah S. Daar¹, Peter A. Singer¹, Deepa Leah Persad¹, Stig K. Pramming², David R. Matthews³, Robert Beaglehole⁴, Alan Bernstein⁵, Leszek K. Borysiewicz⁶, Stephen Colagiuri⁷, Nirmal Ganguly⁸, Roger I. Glass⁹, Diane T. Finegood¹⁰, Jeffrey Koplan¹¹, Elizabeth G. Nabel¹², George Sarna⁶, Nizal Sarrafzadegan¹³, Richard Smith¹⁴, Derek Yach¹⁵ and John Bell¹⁶

Chronic non-communicable diseases (CNCDs) are reaching epidemic proportions worldwide^{1–3}. These diseases — which include cardiovascular conditions (mainly heart disease and stroke), some cancers, chronic respiratory conditions and type 2 diabetes — affect people of all ages, nationalities and classes.

The conditions cause the greatest global share of death and disability, accounting for around 60% of all deaths worldwide. Some 80% of chronic-disease deaths occur in low- and middle-income countries. They account for 44% of premature deaths worldwide. The number of deaths from these diseases is double the number of deaths that result from a combination of infectious diseases (including HIV/AIDS, tuberculosis and malaria), maternal and perinatal conditions, and nutritional deficiencies.

Over the coming decades the burden from CNCDs is projected to rise particularly fast in the developing world. Without concerted action some 388 million people worldwide will die of one or more CNCDs in the next 10 years. With concerted action, we can avert at least 36 million premature deaths by 2015. Some 17 million of these prevented deaths would be among people under the age of 70 (ref. 2).

CNCDs have a huge negative economic impact⁴. In the next 10 years, China, India and the United Kingdom are projected to lose \$558 billion, \$237 billion and \$33 billion, respectively, in national income as a result of heart disease, stroke and diabetes, partly as a result of reduced economic productivity².

Several factors are implicated in this increasing burden, including longer average lifespan, tobacco use, decreasing physical activity, and increasing consumption of unhealthy foods. Fortunately, CNCDs are largely preventable⁵. Up to 80% of premature deaths from heart disease, stroke and diabetes can be averted



Poor diet and smoking are two factors that contribute to the millions of preventable deaths that occur each year.

with known behavioural and pharmaceutical interventions².

Yet the prevention of disability and death from CNCDs gets scant attention worldwide. In sub-Saharan Africa it is understandable that governments, donors and research-funding agencies have channelled most resources into infectious diseases: 5.9% of adults between the ages of 15 and 49 are HIV positive⁶ and malaria alone kills a million children under the age of five each year⁷. In most richer countries the focus of biomedical research on CNCDs has been on treatment rather than prevention.

A crucial aspect of establishing programmes for disease control globally is to identify priorities. To galvanize the health, science and public-policy communities into action on this epidemic, we present here an inventory of 20 grand challenges, grouped under 6 goals, arrived at through a global, structured consensus process.

Two previous 'grand challenge' exercises — the historical one by David Hilbert⁸ in mathematics more than a century ago, and the 2003 Grand Challenges in Global Health initiative spearheaded by the Bill & Melinda Gates Foundation⁹ — showed that the approach focuses significant new attention on an area of study, energizes communities to rise to

meet the challenges, and brings new talent to the field. Although there has been interest in CNCDs among governments in developed countries, research-funding agencies and others¹⁰, this has been incremental and rare in developing nations.

The Grand Challenges in CNCDs we describe here are intended to reduce the global epidemic of these diseases by making the case for worldwide debate, support and funding, and by guiding policy and research in an evidence-based manner.

The Delphi method

To develop the grand challenges, we used the Delphi method — the structured, sequential questioning of a panel, with controlled feedback^{11–14} — to distil knowledge and build reliable consensus among 155 geographically and culturally diverse stakeholders, from 50 countries. We used the following definitions.

A 'grand challenge' was defined as a specific critical barrier that if removed would help to solve an important health problem. The intervention(s) it could lead to might be innovative and, if successfully implemented, would have a high likelihood of impact and feasibility.

'Chronic non-communicable diseases' were defined as diseases or conditions that occur

GRAND CHALLENGES IN CHRONIC NON-COMMUNICABLE DISEASES		
	Grand challenges	Research needed to address goals
Goal A Raise public awareness	<ol style="list-style-type: none"> 1 Raise the political priority of non-communicable disease 2 Promote healthy lifestyle and consumption choices through effective education and public engagement 3 Package compelling and valid information to foster widespread, sustained and accurate media coverage and thereby improve awareness of economic, social and public health impacts 	<ul style="list-style-type: none"> • Study how to engage governments in partnerships for disease prevention • Develop research activities for health that bridge government departments (for example, transport, civic planning, health, education and environment) • Identify reasons for low awareness and advocacy of chronic disease in societies • Study how to create public forums that sustainably raise awareness of issues relating to chronic non-communicable diseases
Goal B Enhance economic, legal and environmental policies	<ol style="list-style-type: none"> 4 Study and address the impact of government spending and taxation on health 5 Develop and implement local, national and international policies and trade agreements, including regulatory restraints, to discourage the consumption of alcohol, tobacco and unhealthy foods 6 Study and address the impacts of poor health on economic output and productivity 	<ul style="list-style-type: none"> • Evaluate the health impacts of agricultural policy interventions • Study the health and economic impacts of comprehensive community-based interventions • Create general population metrics and outcome indicators for policy and programme surveillance. • Quantify impact of chronic non-communicable diseases on domestic economies • Study the international ramifications of changes in food and tobacco consumption • Probe motivations behind domestic expenditures, and how these affect lifestyle choices • Investigate the impact and effectiveness of food-labelling legislation
Goal C Modify risk factors	<ol style="list-style-type: none"> 7 Deploy universally measures proven to reduce tobacco use and boost resources to implement the WHO Framework Convention on Tobacco Control 8 Increase the availability and consumption of healthy food 9 Promote lifelong physical activity 10 Better understand environmental and cultural factors that change behaviour 	<ul style="list-style-type: none"> • Do prospective cohort studies to identify risk factors, the magnitude of their effects, and the factors that reduce risk in chronic non-communicable diseases • Evaluate fetal and early-life influences on chronic disease risk • Find and evaluate new or combined medical preparations to prevent cardiovascular disease and diabetes or reduce their morbidities • Evaluate behavioural modifications to reduce risks • Establish metrics, and relationships between metrics, that are culturally and ethnically specific • Investigate cultural and ethnic variation in risk factors to refine behavioural interventions • Quantify personal risk related to phenotypes, genotypes and multiplicative risks • Study the interaction of environment and genes in risk factors and in outcomes • Develop new biomarkers and diagnostics for risk and for early disease detection
Goal D Engage businesses and community	<ol style="list-style-type: none"> 11 Make business a key partner in promoting health and preventing disease 12 Develop and monitor codes of responsible conduct with the food, beverage and restaurant industries 13 Empower community resources such as voluntary and faith-based organizations 	<ul style="list-style-type: none"> • Study marketing techniques and marketing data derived from commercial companies on behaviour modification • Investigate mechanisms for consumers and the public to exert a positive influence on the food industry • Research the impact of taste, flavour, packaging, labelling and advertising on choice and health • Create and evaluate community-based strategies to promote healthy living • Identify modes of effective public-private partnerships that support health • Develop better understanding of nutrient benefit in foods
Goal E Mitigate health impacts of poverty and urbanization	<ol style="list-style-type: none"> 14 Study and address how poverty increases risk factors 15 Study and address the links between the built environment, urbanization and chronic non-communicable disease 	<ul style="list-style-type: none"> • Investigate the biological basis of health risks related to poverty • Examine the influence of poverty on the adoption of high-risk behaviour • Identify negative effects of economic growth on health • Study how to work with planners, architects and city representatives to enhance the environment for healthier living
Goal F Reorientate health systems	<ol style="list-style-type: none"> 16 Allocate resources within health systems based on burden of disease 17 Move health professional training and practice towards prevention 18 Increase number and skills of professionals who prevent, treat and manage chronic non-communicable diseases, especially in developing countries 19 Build health systems that integrate screening and prevention within health delivery 20 Increase access to medications to prevent complications of chronic non-communicable disease 	<ul style="list-style-type: none"> • Develop strategies to integrate health-system management of communicable and non-communicable disease • Form collaborations to find best practices in delivering affordable and equitable health care • Study how to provide more structured knowledge for health promotion • Develop strategies to ensure that medical training and curricula focus on chronic non-communicable diseases • Develop and provide culturally specific and nationally appropriate resources for training health-care workers • Study how best to ensure that disadvantaged communities have adequate resource allocations in health care and in preventative practice • Optimize use of electronic health records for predicting disease and measuring the effect of health interventions • Study how best to develop and establish real-time surveillance tools • Discover and develop tools for screening and stratifying populations according to risk

in, or are known to affect, individuals over an extensive period of time and for which there are no known causative agents that are transmitted from one affected individual to another. For the purpose of this study the major focus was on cardiovascular diseases, type 2 diabetes, chronic respiratory diseases and certain

cancers. Commonly known risk factors for these include lack of exercise, improper diet and smoking. Note that we excluded mental health and chronic neurological conditions because their risk factors and interventions are so different².

We asked the panel: “What do you think are

the grand challenges in chronic non-communicable diseases?” The first of the three Delphi rounds elicited 1,854 ideas, many of which overlapped. We distilled these into 109 from which the panel selected, ranked and commented on its top 30. These comments and rankings structured the final round: panellists

were asked either to accept the list or to re-rank the choices.

The executive committee and scientific board refined the wording and presentation of the panel's conclusions into the Table published on the previous page (the order does not denote relative importance). Also summarized in the Table are key research needs that the committee and board matched to the goals.

Blueprint for change

The integration of science, technology, policy and social sciences makes the Grand Challenges in CNCDs a particularly comprehensive — if demanding — blueprint for change. For each group of grand challenges we suggest the research now needed.

Many of the grand challenges relate primarily to policy interventions, such as reform of professional training and modification of health systems. These, too, will need evaluation during implementation and ideally should have an evidence base that links research to policy. For example we know about preventing heart attacks and strokes associated with smoking or high blood pressure, but how should we best put these ideas into practice, especially in low-resource settings or on a large scale?

It is not possible to provide a complete list of the research activities that will be required for these grand challenges to be solved systematically, but here we highlight the important ones. Interdisciplinary research will be needed, for example, to explore the interactions of behaviour, environment and genetics in framing risks and determining outcomes. Such research will also need to focus on equity, and on the effects of gender and culture on risk, the effectiveness of interventions and access to health care. Ethical, social, cultural and sustainability issues must be addressed before emerging interventions and technologies can be taken up by communities and incorporated into public-health and health-care systems^{15,16}. Data and research repositories will also be essential, and standardization, where possible, will allow international comparisons and help global partnerships.

Although these challenges are applicable to all countries, different nations should identify local priorities from among those identified here for immediate attention, depending on resources and disease patterns.

Next steps

Addressing the challenges identified here requires the participation of governments, the World Health Organization, the World Bank, regional development banks, foundations, research-funding agencies, donor agencies and others. The business community and civil society organizations will also be crucial partners. A global governance mechanism to coordinate this work across different sectors will be important to prevent dissipation or duplication of effort.

With this publication, the Grand Challenges Global Partnership is being established with a secretariat in the Oxford Health Alliance (OxHA). It will be funded for the first 5 years by members of OxHA (www.oxha.org). The partnership is intended primarily as a coordinating body for research-funding organizations and to harmonize efforts among other relevant initiatives^{10,17}. The founding partners, OxHA, the UK Medical Research Council, the Canadian Institutes of Health Research, the Indian Council of Medical Research and the US National Institutes of Health intend to expand the partnership, forge collaborative research opportunities, and monitor progress towards meeting these challenges. An advocacy programme will also be developed, to encourage adoption of the challenges and goals.

With the Grand Challenges in Global Health (www.gcgh.org) initiative there was an upfront commitment of US\$200 million (later increased to \$450 million) in research funding. By contrast, the Grand Challenges in CNCDs are not linked directly to a funding programme. These problems require initial financing, a long-term commitment and a coordinated effort between multiple funding agencies around a set of clear priorities. Providing such priorities is the major goal of this grand-challenge exercise. The growing interest in this area of research now being registered by governments and funding agencies suggests that substantial resources may be available in the future.

In the first instance, the main function of the Grand Challenges Global Partnership will be monitoring and reporting. It will provide cross-referencing between agencies to ensure efforts are complementary and that major objectives are not overlooked. We will therefore prepare for the research-funding agencies and foundations an annual progress report, beginning a year from now.

Chronic non-communicable diseases constitute the major burdens of illness and disability in almost all countries of the world. They must urgently receive more resources, research and attention, as mapped out in these grand challenges. Inaction is costing millions of premature deaths throughout the world. ■

1. Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T. & Murray, C. J. L. (eds) *Global Burden of Disease and Risk Factors* (Oxford Univ. Press and World Bank, Washington DC, 2006).
2. World Health Organization *Preventing Chronic Diseases: A Vital Investment* (WHO, Geneva, 2005).
3. Adeyi, O., Smith, O. & Robles, S. *Public Policy and the Challenge of Chronic Noncommunicable Diseases* (World Bank, Washington DC, 2007).
4. Suhrcke, M., Nugent, R. A., Stuckler, D., & Rocco, L. *Chronic Disease: An Economic Perspective* (Oxford Health Alliance, London, 2006).
5. Jamison, D. T. et al. (eds) *Priorities in Health* (World Bank, Washington DC, 2006); available at <http://www.dcp2.org>
6. UNAIDS *AIDS Epidemic Update: Special Report on HIV/AIDS December 2006* (UNAIDS/WHO, Geneva, 2006); available at http://data.unaids.org/pub/EpiReport/2006/2006_EpiUpdate_en.pdf
7. Breman, J. G., Alilio, M. S. & Mills, A. (eds) *Am. J. Trop. Med. Hyg.* **71** (suppl. 2), 1–282 (2004).

8. Hilbert, D. *Bull. Am. Math. Soc.* **8**, 437–479 (1902).
9. Varmus, H. et al. *Science* **302**, 398–399 (2003).
10. Yach, D. & Prammings, S. *Bulletin of Medicus Mundi Switzerland* No. 106, 12–15 (November 2007).
11. Jones, J. & Hunter, D. *Br. Med. J.* **311**, 376–380 (1995).
12. Martin, B. R. *Technology Analysis & Strategic Management* **7**, 139–168 (1995).
13. Powell, C. *Journal of Advanced Nursing* **41**, 376–382 (2003).
14. Rowe, G. & Wright, G. in *Principles of Forecasting* (ed. Armstrong, J. S.) 125–144 (Kluwer Academic Publishers, Norwell, MA, 2001).
15. Singer, P. A. et al. *Nature* **449**, 160–163 (2007).
16. Berndtson, K. et al. *PLoS Med.* **4**, e268 (2007).
17. *Lancet* **370**, 630 (2007).

Supplementary information See www.mrcglobal.org/supplemental/nature_grandchallenges or www.OxHA.org for supplementary materials, including details of methodology, results, the Delphi study panel, and membership of the executive committee and scientific board.

Acknowledgements The study was carried out at the McLaughlin-Rotman Centre for Global Health (University Health Network/University of Toronto). It was funded by the Oxford Health Alliance, with contributions from the UK Medical Research Council and the McLaughlin-Rotman Centre for Global Health. We are grateful for valuable advice from Harold Varmus. We thank Gunjeet Minhas, Dilnoor Panjwani, Mark Messih, Munira Tayabali and Marisa Pulaski for their valuable assistance with this project.

Author contributions A.S.D. and D.L.P. carried out the Delphi study. J.B., A.B., R.B., L.K.B., A.S.D., R.I.G. and J.K. are members of the executive committee. S.C., N.G., D.T.F., D.R.M., S.K.P., G.S., N.S., R.S., P.A.S. and D.Y. are members of the scientific board. They all contributed to the study. All authors contributed to the writing of the paper.

Author declaration The authors declare no competing financial interests.

Author addresses ¹Program on Life Sciences, Ethics and Policy, McLaughlin-Rotman Centre for Global Health, University Health Network/University of Toronto, 101 College Street, Toronto, Ontario, M5G 1L7, Canada. ²The Oxford Health Alliance, 1st Floor, 28 Margaret Street, London W1W 8RZ, UK. ³Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Headington, Oxford, OX3 7LJ, UK. ⁴Faculty of Medical and Health Sciences, University of Auckland, 85 Park Road, Grafton, Auckland 1142, New Zealand. ⁵Canadian Institutes of Health Research, 160 Elgin Street, 9th Floor, Address Locator 4809A, Ottawa, Ontario, K1A 0W9, Canada. ⁶Medical Research Council, 20 Park Crescent, London W1B 1AL, UK. ⁷Diabetes Centre and Department of Endocrinology, Metabolism and Diabetes, Prince of Wales Hospital, University of Sydney, Baker Street, Randwick, New South Wales 2031, Australia. ⁸Indian Council of Medical Research, Ansari Nagar, New Delhi, 110029, India. ⁹Fogarty International Center, National Institutes of Health, Building 31, Room B2C02, 31 Center Drive, MSC 2220, Bethesda, Maryland 20892-2220, USA. ¹⁰Canadian Institutes of Health Research, Institute of Nutrition, Metabolism and Diabetes, 160 Elgin Street, 9th Floor, Address Locator 4809A, Ottawa, Ontario, K1A 0W9, Canada. ¹¹Global Health Institute, Emory University, 1440 Clifton Road, Atlanta, Georgia 30322, USA. ¹²National Heart, Lung, and Blood Institute, National Institutes of Health, Building 31, Room 5A48, 31 Center Drive, MSC 2486, Bethesda, Maryland 20892-2486, USA. ¹³Isfahan Cardiovascular Research Center, Isfahan University of Medical Sciences, Seddigheh Tahereh Research and Treatment Hospital, Khorram Avenue, Isfahan, Iran. ¹⁴Ovations Chronic Disease Initiative, UnitedHealth Europe, London SW1P 1SB, UK. ¹⁵PepsiCo, 700 Anderson Hill Road, Purchase, New York 10577, USA. ¹⁶University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK.

Genotyping of *Rhesus* SCNT pluripotent stem cell lines

Arising from: J. A. Byrne *et al.* *Nature* 450, doi:10.1038/nature06357 (2007)

Somatic cell nuclear transfer (SCNT) into enucleated oocytes has emerged as a technique that can be used to derive mouse embryonic stem cell lines with defined genotypes. In this issue Byrne *et al.*¹ report the derivation of two SCNT *Rhesus macaca* male stem cell lines designated CRES-1 and CRES-2. Molecular studies detailed in their paper provides supporting evidence that the chromosome complement of CRES-1 and CRES-2 was genetically identical to the male cell donor nucleus and that the mitochondrial DNA originated from different recipient oocytes. In this validation paper, we independently confirm that both stem cell lines were indeed derived by SCNT.

The aim of our validation experiments was to confirm the findings of Byrne *et al.*¹, and to exclude any possibility that the CRES-1 and CRES-2 cell lines could have been derived from *in vitro* fertilization (IVF) blastocysts or those formed from parthenogenetic oocyte activation. IVF-derived stem cells are identifiable by a genetic profile comprising half the maternal chromosome complement, whereas parthenogenetically derived stem cell lines show a complete maternal chromosome profile, although there is a tendency for homozygosity, particularly towards the centromeric regions that reflect recombination signatures².

Independent DNA samples from the two female oocyte recipients, the male cell nucleus donor and the two stem cell lines CRES-1 and CRES-2 were provided by the Oregon National Primate Research Center. Samples were collected by T. Hobbs, who observed the DNA extraction procedure in S. M. Mitalipov's laboratory. The five

DNA samples were coded and sent to M. Pera's laboratory at the Center for Stem Cell and Regenerative Medicine, University of Southern California, where we performed the genetic validation studies.

The analysis of CRES-1 and CRES-2 outlined in Byrne *et al.*¹ used homologous human STR markers for genotyping, including 25 STRs located on 19 of the 20 *Rhesus* autosomes, 14 MHC-linked STR markers, 30 informative single nucleotide polymorphisms (SNPs) and the amelogenin marker for sex determination. For independent genotyping, we selected ten informative human STR markers and amelogenin, 54 recently mapped *Rhesus macaque* STR markers (average heterozygosity of 0.73), distributed across all 20 autosomes³, and five human Y-chromosome-specific STR markers⁴. In addition, we undertook a more comprehensive analysis of *Rhesus* chromosomes 1 and 3, using 22 and 16 homologous human STR markers⁵, respectively, distributed from centromeric to telomeric regions. After STR analysis, samples were decoded by T. Hobbs. The CRES-1 and CRES-2 allelic profiles for amelogenin and 82 STR markers (20 of 102 STR markers did not amplify across all samples) matched completely with the male somatic donor nucleus (Fig. 1a, and Table 1 of the Appendix). Furthermore, the five Y-chromosome markers that were positive on human male genomic DNA were negative on all five samples, supporting the conclusion that both cell lines were of *Rhesus* origin.

On the basis of three informative SNPs within the polymorphic D-loop region of mitochondrial DNA, Byrne *et al.*¹ also provided

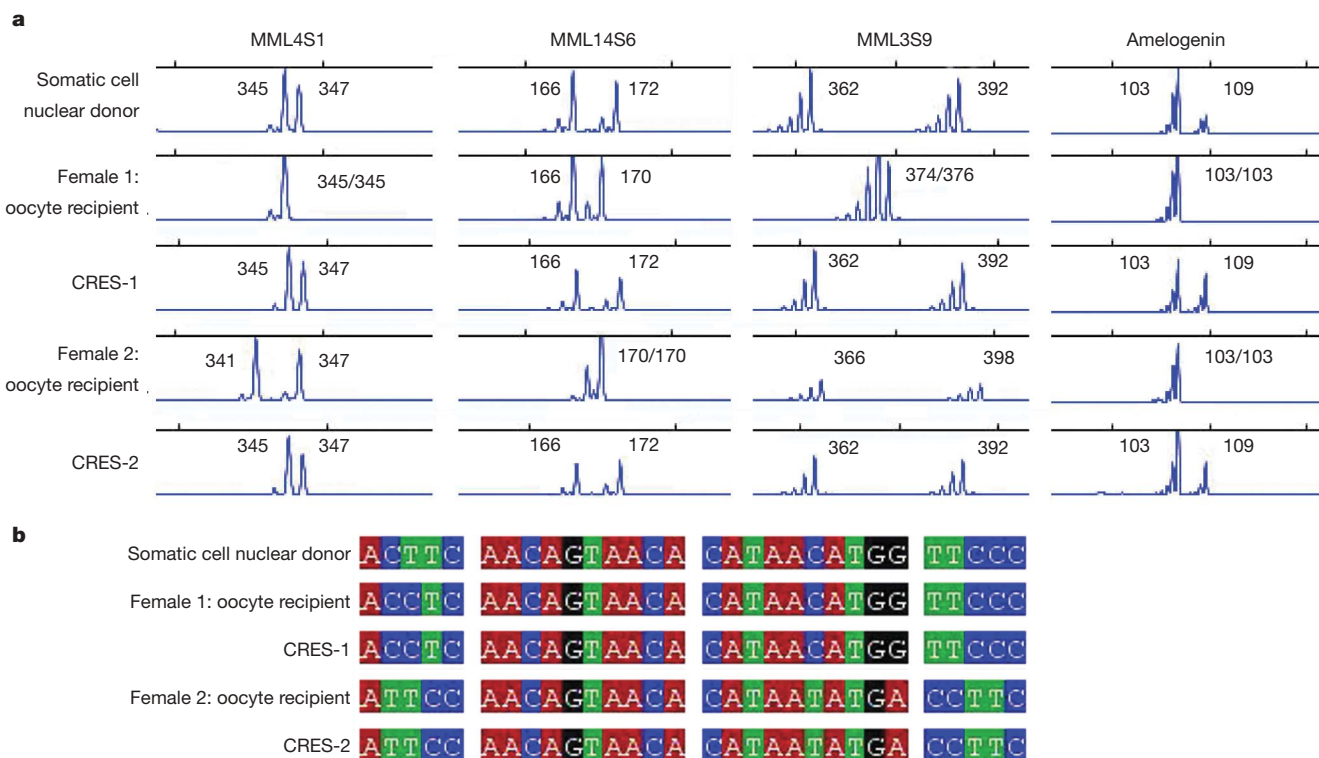


Figure 1 | Analysis of nuclear and mitochondrial DNA origin in CRES-1 and CRES-2 stem cell lines by STR genotyping and DNA sequencing. a, Profiles of three (MML4S1, MML14S6 and MML3S9) of 54 *Rhesus macaque* STR markers and amelogenin are shown as examples. Blue peaks (6-FAM labelled PCR products) represent STR alleles and numbers signify allele sizes in nucleotides. The STR profiles of the male donor cell match with that of

CRES-1 and CRES-2 cell lines. **b**, DNA sequence comparison of the mitochondrial D-loop region. The most polymorphic region (position 16212–16241 of the *Macaca mulatta* NCBI reference sequence NC_005943) is shown as an example. Colour coding confirms mitochondrial inheritance between CRES-1/female oocyte recipient 1 and CRES-2/female oocyte recipient 2.

evidence that CRES-1 and CRES-2 were derived from their respective female recipient oocytes. To confirm mitochondrial inheritance, we performed extensive DNA sequencing across the D-loop hypervariable region (nucleotides 403–718 and 15957–16506 of the *Macaca mulatta* NCBI reference sequence NC_005943) and identified 83 nucleotide variants that definitively genotyped each mitochondrial DNA sample. Complete D-loop sequence matches were identified for CRES-1/female oocyte recipient 1 and for CRES-2/female oocyte recipient 2 that differed from those of the male donor at 5 and 79 nucleotide positions, respectively (Fig. 1b, and Table 2 of the Appendix). There was no evidence in the CRES-1 or CRES-2 cell lines for heterozygosity at any of the 83 nucleotide variants that would indicate any low levels of surviving mitochondria originating from the male donor cell.

Taken together, independent genotyping of nuclear and mitochondrial DNA in the five *Rhesus* samples demonstrates beyond any doubt that CRES-1 and CRES-2 were indeed derived from SCNT primate blastocysts and not fertilized or parthenogenetic embryos. Thus, proof of concept for creating SCNT primate stem cells is now firmly established.

METHODS

Independent DNA samples were obtained from female oocyte recipients 1 and 2 (20733 and 19735), the male donor (19822) and the CRES-1 and CRES-2 stem cell lines. *Rhesus macaque*² and human^{3,4} chromosomal STR markers and corresponding primer pairs have been described. Genomic DNA (30 ng) was amplified by 35 cycles of fluorescent polymerase chain reaction (PCR) using the Platinum Taq DNA Polymerase system (Invitrogen). STR PCR products were separated on an ABI 3730 DNA Analyser and alleles were sized using Peak Scanner Software. Mitochondrial D-loop PCR products were directly sequenced using the Big Dye cycle sequencing method. Genescan analysis and DNA sequencing was performed at the University of Southern California DNA Core Facility.

David S. Cram¹, Bi Song¹ & Alan O. Trounson¹

¹Monash Immunology and Stem Cell Laboratories, Monash University, Clayton 3800, Victoria, Australia.

e-mail: david.cram@med.monash.edu.au

Received 9 October; accepted 6 November 2007.

- Byrne, J. A. *et al.* Producing primate embryonic stem cells by somatic cell nuclear transfer. *Nature* (this issue) doi:10.1038/nature06357 (2007).
- Kim, K. *et al.* Histocompatible embryonic stem cells by parthenogenesis. *Science* **315**, 482–486 (2007).
- Raveendran, M. *et al.* Designing new microsatellite markers for linkage and population genetic analyses in rhesus macaques and other nonhuman primates. *Genomics* **88**, 706–710 (2006).
- Budowle, B. *et al.* Twelve short tandem repeat loci Y chromosome haplotypes: genetic analysis on populations residing in North America. *Forensic Sci. Int.* **150**, 1–15 (2005).
- Rogers, J. *et al.* An initial genetic linkage map of the rhesus macaque (*Macaca mulatta*) genome using human microsatellite loci. *Genomics* **87**, 30–38 (2006).

doi:10.1038/nature06456

APPENDIX

This Appendix contains two tables, showing the genotyping and sequencing results. The data confirm the expected nuclear and mitochondrial DNA inheritance patterns for the two SCNT-derived *Rhesus* stem cell lines CRES-1 and CRES-2.

Table 1 | Nuclear DNA inheritance pattern of CRES-1 and CRES-2 cell lines

STR marker	Female 1 oocyte donor for CRES-1	Female 2 oocyte donor for CRES-2	Male nuclear donor for CRES- 1/2	CRES-1 cell line	CRES-2 cell line
D3S1768	216/216	202/210	190/214	190/214	190/214
D7S513	189/197	201/201	TAF	185/185	185/185
D11S925	223/249	223/231	223/223	223/223	223/223
D13S765	224/236	208/214	224/252	224/252	224/252
D15S823	324/338	320/342	318/352	318/352	318/352
D17S1300	228/274	238/274	266/270	266/270	266/270
D22S685	306/310	282/294	306/318	306/318	306/318
D6S276	202/202	196/196	206/214	206/214	206/214
D6S1691	192/198	196/196	194/194	TAF	TAF
G51152	192/206	212/212	192/216	192/216	192/216
Amelogenin	103/103	103/103	103/109	103/109	103/109

Table 1 | Continued

STR marker	Female 1 oocyte donor for CRES-1	Female 2 oocyte donor for CRES-2	Male nuclear donor for CRES- 1/2	CRES-1 cell line	CRES-2 cell line
DYS391	TAF	TAF	TAF	TAF	TAF
DYS392	TAF	TAF	TAF	TAF	TAF
DYS437	TAF	TAF	TAF	TAF	TAF
DYS438	TAF	TAF	TAF	TAF	TAF
DYS439	TAF	TAF	TAF	TAF	TAF
D1S2644	222/222	222/222	218/228	218/228	218/228
D1S2864	134/148	148/166	134/152	134/152	134/152
D1S496	156/156	154/160	160/160	160/160	TAF
D1S231	150/164	152/164	150/164	150/164	150/164
D1S192	200/200	206/214	190/206	190/206	190/206
D1S207	TAF	TAF	TAF	TAF	TAF
D1S2896	168/170	170/170	170/172	170/172	170/172
D1S248	210/210	210/210	210/210	210/210	210/210
D1S534	170/172	170/172	172/172	172/172	172/172
D1S2635	127/127	119/123	119/125	119/125	119/125
D1S426	144/144	144/144	144/148	144/148	144/148
D1S2878	220/220	220/220	216/220	216/220	216/220
D1S194	236/236	236/238	228/236	228/236	228/236
D1S431	120/120	124/126	120/120	120/120	120/120
D1S215	180/180	186/198	190/198	190/198	190/198
D1S518	226/226	230/242	218/222	218/222	218/222
D1S533	180/180	184/212	180/208	180/208	180/208
D1S2685	184/186	190/196	184/186	184/186	184/186
D1S213	102/102	94/102	98/102	98/102	98/102
D1S1656	140/148	132/144	140/140	140/140	140/140
D1S437	328/328	246/326	268/328	TAF	TAF
D1S1594	148/152	156/156	144/156	144/156	144/156
D21S1259	202/208	198/198	196/200	196/200	196/200
D21S1887	138/140	134/138	140/140	140/140	140/140
D21S1246	412/432	388/428	420/424	420/424	420/424
D21S2054	176/180	180/184	158/172	158/172	158/172
D21S2053	224/232	228/236	224/228	224/228	224/228
D21S1904	TAF	TAF	TAF	TAF	TAF
D7S493	206/218	206/206	206/212	206/212	206/212
D7S2534	TAF	TAF	TAF	TAF	TAF
D7S555	104/104	TAF	104/104	104/104	104/104
D7S2550	258/262	250/258	258/262	258/262	258/262
D7S524	182/196	192/196	190/190	190/190	190/190
D7S496	120/124	120/120	112/118	112/118	112/118
D7S2438	230/246	222/222	222/232	222/232	222/232
D7S794	TAF	TAF	TAF	TAF	TAF
D7S1826	116/116	116/116	116/116	116/116	116/116
D7S559	186/186	186/194	172/172	172/172	172/172
MML1S1	283/283	283/289	279/283	279/283	279/283
MML1S5	125/153	125/153	111/153	111/153	111/153
MML1S8	275/275	275/275	275/275	275/275	275/275
MML1S10	151/151	123/149	149/149	149/149	149/149
MML1S13	239/249	245/245	245/249	245/249	245/249
MML2S2	226/244	224/232	216/224	216/224	216/224
MML3S4	234/248	234/238	232/248	232/248	232/248
MML3S7	246/258	250/258	246/260	246/260	246/260
MML3S9	374/376	366/398	362/392	362/392	362/392
MML3S13	236/236	226/226	TAF	234/236	234/236
MML3S16	104/108	106/108	TAF	104/106	104/106
MML3S29	446/446	448/448	436/450	436/450	436/450
MML4S1	345/345	341/347	345/347	345/347	345/347
MML4S8	288/300	294/300	288/294	288/294	288/294
MML5S3	323/323	313/321	323/323	323/323	323/323
MML5S6	222/222	212/222	212/212	212/212	212/212
MML5S9	208/220	204/216	198/212	198/212	198/212
MML6S3	162/162	152/162	162/162	162/162	162/162
MML6S5	TAF	147/155	149/149	149/149	149/149
MML6S8	330/330	324/330	324/328	324/328	324/328
MML6S12	214/214	206/212	206/214	206/214	206/214
MML7S2	400/402	398/400	400/408	400/408	400/408
MML7S4	156/156	158/158	156/156	156/156	TAF
MML7S6	184/190	184/188	184/184	184/184	184/184
MML7S7	305/305	305/305	305/305	305/305	305/305
MML8S3	288/294	288/296	294/294	294/294	294/294
MML8S7	158/162	152/158	150/158	150/158	150/158
MML9S1	148/160	148/160	142/164	142/164	142/164
MML9S6	396/404	404/406	396/404	396/404	396/404
MML9S8	164/164	160/168	150/168	150/168	TAF

Table 1 | Continued

STR marker	Female 1 oocyte donor for CRES-1	Female 2 oocyte donor for CRES-2	Male nuclear donor for CRES- 1/2	CRES-1 cell line	CRES-2 cell line
MML9S10	298/306	296/300	300/310	300/310	300/310
MML9S18	208/214	206/212	218/226	218/226	218/226
MML11S1	138/150	142/150	136/150	136/150	136/150
MML11S2	160/172	160/174	174/174	174/174	174/174
MML11S16	228/238	228/232	238/238	238/238	238/238
MML14S2	143/143	139/143	139/139	139/139	139/139
MML14S6	166/170	170/170	166/172	166/172	166/172
MML14S7	174/174	182/182	174/180	174/180	174/180
MML14S10	224/230	224/230	218/218	218/218	218/218
MML14S15	444/458	452/456	428/462	428/462	428/462
MML15S3	208/208	228/232	232/238	232/238	232/238
MML15S5	214/214	214/214	214/214	214/214	TAF
MML15S7	146/146	146/148	144/146	144/146	144/146
MML16S1	TAF	390/394	390/390	390/390	390/390
MML16S3	133/147	137/145	147/147	147/147	147/147
MML16S5	351/351	351/351	351/351	351/351	TAF
MML16S10	147/161	143/175	143/161	143/161	TAF
MML17S1	322/322	322/326	326/330	326/330	326/330
MML17S2	161/163	165/167	161/161	161/161	161/161
MML17S7	185/185	185/185	181/181	181/181	TAF
MML17S8	TAF	TAF	TAF	TAF	TAF
MML17S11	240/240	232/240	230/234	230/234	230/234
MML18S3	212/212	212/218	210/218	210/218	210/218
MML19S1	334/334	336/336	334/336	334/336	334/336

Codes 'D' and 'MML' are prefixes used for human and *Rhesus* STR markers, respectively. D1S STR markers are homologous to *Rhesus* chromosome 1. D21S and D7S STR markers are homologous to *Rhesus* chromosome 3. Numbers correspond to allelic sizes in nucleotides, calculated by Genescan analysis. TAF, total amplification failure.

Table 2 | Mitochondrial DNA inheritance pattern of CRES-1 and CRES-2 cell lines

Nucleotide position in mitochondrial reference sequence	Female 1 oocyte donor for CRES-1	Female 2 oocyte donor for CRES-2	Male nuclear donor for CRES-1/2	CRES-1 cell line	CRES-2 cell line
407–410	GCAC	AACA	GCAA	GCAC	AACA
417	deletion A	deletion A	A	deletion A	deletion A
425	C	T	C	C	T
438	T	C	T	T	C
454	T	C	T	T	C
472	G	A	A	G	A
478	A	G	A	A	G
500	C	T	C	C	T
507	G	A	G	G	A
511–512	GC	AT	GC	GC	AT
520	C	T	C	C	T
523	T	C	T	T	C
527–528	AG	GA	AG	AG	GA

Table 2 | Continued

Nucleotide position in mitochondrial reference sequence	Female 1 oocyte donor for CRES-1	Female 2 oocyte donor for CRES-2	Male nuclear donor for CRES-1/2	CRES-1 cell line	CRES-2 cell line
534	T	C	T	T	C
589	T	C	T	T	C
642	G	A	G	G	A
661	A	G	A	A	G
689	C	T	C	C	T
697	C	T	C	C	T
15999	T	C	T	T	C
16024	A	G	A	A	G
16028	C	T	C	C	T
16035	T	C	T	T	C
16040	-	insertion A	-	-	insertion A
16047	T	C	T	T	C
16058	T	C	T	T	C
16060	C	T	C	C	T
16063	C	T	C	C	T
16071	C	T	C	C	T
16153	T	C	T	T	C
16162	C	T	C	C	T
16167–16169	TTG	CCA	TTG	TTG	CCA
16176–16178	CAA	TGG	CAA	CAA	TGG
16180	C	T	C	C	T
16182	G	A	G	G	A
16201–16202	TC	CT	TC	TC	CT
16209	C	T	C	C	T
16213–16215	CCT	TTC	CTT	CCT	TTC
16232	C	T	C	C	T
16236–16240	GTTCC	ACCTT	GTTCC	GTTCC	ACCTT
16244–16245	GG	AA	GG	GG	AA
16247	C	T	C	C	T
16250	A	G	A	A	G
16252–16254	CTT	TCC	CTT	CTT	TCC
16257	C	deletion C	C	C	deletion C
16260	C	T	C	C	T
16267–16268	GA	AG	GA	GA	AG
16272	T	C	T	T	C
16277–16278	TA	CG	TA	TA	CG
16288	T	A	T	T	A
16295	T	C	T	T	C
16313	G	A	G	G	A
16338–16339	CT	TC	CT	CT	TC
16349	C	T	T	C	T
16368	T	C	T	T	C
16373	T	C	T	T	C
16384	G	A	G	G	A
16391–16392	AC	GT	AC	AC	GT
16477	C	T	C	C	T
16482	C	T	C	C	T

Hypervariable D-loop regions (nucleotide positions 403–718 and 15957–16506) of the *Macaca mulatta* NCBI reference sequence NC_005943 were analysed. The 83 positions of nucleotide variation (substitution, deletion or insertion) are listed.

Evidence for mantle plumes?

Arising from: B. Bourdon, N. M. Ribe, A. Stracke, A. E. Saal & S. P. Turner *Nature* **444**, 713–717 (2006)

Geophysical hotspots have been attributed to partially molten asthenosphere, fertile blobs, small-scale convection and upwellings driven by core heat^{1–4}. Most are short-lived or too close together to be deeply seated, and do not have anomalous heat flow^{5,6} or temperature^{7,8}; many are related to tectonic features^{9–11}. Bourdon *et al.*¹² investigate the dynamics of mantle plumes from uranium-series geochemistry and interpret their results as evidence for thermal plumes. Here we show why alternative mechanisms of upwelling and melting should be considered.

Revised estimates of temperature, heat flow and buoyancy at ridges and hotspots, and developments in plume^{3,13} and plate theory^{2,9,10} are relevant to the conclusions of Bourdon *et al.*¹². No near-ridge hotspot has anomalous temperature⁷ and no hotspot has a significant heat flow anomaly^{5,6}. The only active hotspot with a petrology-based temperature higher than mid-ocean-ridge basalt is, arguably, Hawaii¹⁴. Hawaiian lithosphere, however, is 140 °C colder than predicted by plume models¹⁵. It is no longer generally argued that all, or even many, hotspots are due to deep plumes, but lower mantle conditions are considered by many investigators as essential to the rationalization of observations¹². However, tectonics and shallow low-melting heterogeneities, rather than excess temperatures, may be responsible for hotspots^{2,11}.

To satisfy global constraints, narrow plumes must have ascent rates and temperatures much greater than the broad upwellings associated with plate tectonics and normal mantle convection^{3,13}. Required excess temperatures are 200–300 °C and velocities are one-half to tens of metres per year. The absence of evidence for such high values has been rationalized in several ways¹², but, taken at face value, supports a plate tectonic and lithologic, rather than thermal, explanation for hotspots.

Uranium-series geochemistry may provide insight that is independent of previous arguments¹². Bourdon *et al.*¹² interpret U-series model data as independent evidence for thermal plumes and evidence that hotter plumes are stronger. Plumes are modelled as if they originated in a deep thermal boundary layer¹² heated from below, although the data do not constrain anything deeper than about 60–100 km. Shallow processes^{9–11} and homologous temperature (T_H) variations^{2,3} are not considered (plume simulations use a homogeneous mantle); modelling assumptions and parametrizations to date do not permit a shallow or non-thermal interpretation. Shallow chemical buoyancy can mimic effects of temperature, including isotope gradients. Could U-series data discriminate between buoyant decompression melting of shallow fertile blobs^{1,10} and deep thermal plumes? Model velocities are comparable to plate and passive upwelling velocities, and much less than 0.5 m yr^{-1} .

Melt retention buoyancy in low-melting silicates (fertile blobs) is the equivalent of 300–600 °C temperature excess⁴; lowering the solidus or raising the temperature provide equivalent buoyancies³. Fertile blobs absorb mantle heat and turn into buoyant diapirs². High- T_H blobs can melt deeper, rise faster and retain melt longer than subsolidus ambient mantle (low- T_H) rising passively under ridges. Lateral flow of material beneath the lithosphere has been taken as evidence for plumes⁵, but is equally consistent with spreading of chemically buoyant or high- T_H blobs.

Hawaii has conflicting petrological^{8,14}, fertility and tectonic^{10,11} interpretations and cannot be understood in terms of temperature alone. Eruption rates peak at the large-offset 300-km-wide Molokai fracture zone¹¹. A plausible explanation involves underplating, decompression melting of 'eclogite' at lithosphere steps¹⁰ and ascent through Molokai fracture zone conduits. The time between melt extraction and eruption, under these conditions, is unlikely to be similar to ridges.

If hotspots involve tectonic and T_H variations, then conventional fluid dynamic constraints on depth, temperature and velocity are removed. Temperatures of thermal plumes cannot be arbitrarily low^{3,13}. Processes that utilize composition, internal heating, lateral flow, stress and ponding to localize volcanism can operate at lower temperatures^{1,9}. The definition of mantle plume, to be useful, should recognize the difference.

Don L. Anderson¹ & James H. Natland²

¹Seismological Laboratory, California Institute of Technology, MC 252-21, Pasadena, California 91125, USA.

e-mail: dla@gps.caltech.edu

²Marine Geology and Geophysics, Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida 33149, USA.

Received 21 March; accepted 3 October 2007.

1. Anderson, D. L. in *Plates, Plumes, and Paradigms* (eds Foulger, G. R., Natland, J. H., Presnall, D. C. & Anderson, D. L.) 31–54 (GSA Special Paper 388, Geological Society of America, Boulder, 2005).
2. Anderson, D. L. in *Plates, Plumes, and Planetary Processes* (eds Foulger, G. R. & Jurdy, D. M.) 47–64 254–255 (GSA Special Paper 430, Geological Society of America, Boulder, 2007).
3. Sleep, N. H. in *Plates, Plumes, and Planetary Processes* (eds Foulger, G. R. & Jurdy, D. M.) 29–45 (GSA Special Paper 430, Geological Society of America, Boulder, 2007).
4. Yamamoto, M. *et al.* in *Plates, Plumes, and Planetary Processes* (eds Foulger, G. R. & Jurdy, D. M.) 165–188 (GSA Special Paper 430, Geological Society of America, Boulder, 2007).
5. Clift, P. D. in *Plates, Plumes and Paradigms* (eds Foulger, G. R., Natland, J. H., Presnall, D. C. & Anderson, D. L.) 279–287 (GSA Special Paper 388, Geological Society of America, Boulder, 2005).
6. Stein, C. A. & Von Herzen, R. P. in *Plates, Plumes and Planetary Processes* (eds Foulger, G. R. & Jurdy, D. M.) 261–274 (GSA Special Paper 430, Geological Society of America, Boulder, 2007).
7. Presnall, D. C. & Gudfinnsson, G. H. Origin of the oceanic lithosphere. *J. Petrol.* (in the press).
8. Fiallon, T. J. *et al.* in *Plates, Plumes and Planetary Processes* (eds Foulger, G. R. & Jurdy, D. M.) 235–260 (GSA Special Paper 430, Geological Society of America, Boulder, 2007).
9. McNutt, M. K. & Bonneville, A. A shallow, chemical origin for the Marquesas Swell. *Geochim. Geophys. Res.* **5** (6), doi:10.1029/1999GC000028 (2000).
10. Raddick, M. J. *et al.* Buoyant decompression melting: A possible mechanism for intraplate volcanism. *J. Geophys. Res.* **107**, doi:10.1029/2001JB000617 (2002).
11. Winterer, E. L. & Natland, J. H. in *Plates, Plumes, and Planetary Processes* (eds Foulger, G. R. & Jurdy, D. M.) 230–231 (GSA Special Paper 430, Geological Society of America, Boulder, 2007).
12. Bourdon, B. *et al.* Insights into the dynamics of mantle plumes from uranium-series geochemistry. *Nature* **444**, 713–717 (2006).
13. Larson, T. & Yuen, D. Fast plumeheads: Temperature-dependent versus non-Newtonian rheology. *Geophys. Res. Lett.* **24**, 1995–1998 (1997).
14. Herzberg, C. *et al.* Temperatures in ambient mantle and plumes: Constraints from basalts, picrites, and komatiites. *Geochim. Geophys. Res.* **8**, Q02006, doi:10.1029/2006GC001390 (2007).
15. Sen, G. *et al.* Hawaiian mantle xenoliths and magmas: composition and thermal character of the lithosphere. *Am. Mineral.* **90**, 871–887 (2005).

doi:10.1038/nature06376

Bourdon et al. reply

Replying to: D. L. Anderson & J. H. Natland *Nature* **450**, doi:10.1038/nature06376 (2007)

Anderson and Natland's comment¹ does not question our results regarding the velocity structure of mantle upwellings based on uranium-series². We stated clearly that our results do not provide direct measurements of the depth from which hotspot volcanism starts; we have not made unjustified claims by saying we have identified deep mantle plumes². However, our results do shed light on the still incompletely understood causes of hotspot volcanism².

Many responses to the criticism of the plume model¹ have been published^{3,4} and we shall not reiterate these arguments, which refute a low mantle temperature beneath hotspots based on petrology. First, Anderson and Natland argue that buoyancy beneath 'hotspots' could be explained by fertile blobs. In general, buoyancy is driven by contrast in temperature, fertility and melt fraction⁵. As fertility is usually associated with a higher iron content (and hence a greater density), the effect of melting out the fertile (dense) component will not increase the buoyancy relative to ambient mantle. Thus, the only source of extra buoyancy one can consider in the model of Anderson and Natland¹ is buoyancy due to melt retention (or melt fraction ϕ). In this respect, U-series provide clear clues on the melt fraction during melting and all studies show that it should not exceed a few permil (ref. 2). Hence, the effect of a fertile source on the melt fraction (for $\phi = 0.003$) should in fact be limited and be equivalent to an excess temperature of 10 °C. For these reasons, we do not think a fertile blob can be the source of buoyancy beneath hotspots.

Second, if the increased melting rates were due to the presence of fertile blobs, then there should be a correlation between clear indices of enrichment, such as radiogenic isotopes, and U-series activity ratios. For two localities we used (the Galapagos and the Azores), such a correlation does not exist. Furthermore, as fertility is associated with enrichment in water (at least in the Galapagos and the Azores), this will decrease the melting rate of the blob by at least a factor of ten, which could easily compensate the effect of increased fertility on melting rates. These combined effects would fail to explain the U-series observations. We do not believe that fertile blobs can be a general explanation.

Third, although the mantle upwelling rates determined using U-series may not be entirely reliable in absolute terms, when compared with mid-ocean-ridge settings, they clearly indicate faster upwelling near the centre of the plume than beneath adjacent ridges. This is evident in the Azores region^{2,6} (where the spreading rate is slow) but also in the Galapagos (A.S. *et al.*, manuscript in preparation), where the spreading rate is faster. The case of the Azores is particularly interesting because it shows that despite the relatively small buoyancy flux, the mantle is upwelling faster than beneath the nearby Mid-Atlantic Ridge. Furthermore, estimates of mantle

upwelling velocities beneath Hawaii⁷ do show that velocities of up to several metres per year are estimated.

Last, the temperatures of hotspot magmas have been widely discussed (see ref. 8 for a critical review of ridge and hotspot mantle temperatures) and these consistently indicate that temperatures are hotter than that of the ambient mantle, even when the variability of normal potential temperatures of mid-ocean-ridge basalts (120 °C; ref. 7) is taken into account. Although we did not focus on a comparison with ridges², U-series unambiguously indicate higher temperatures beneath hotspots than at mid-ocean ridges⁹.

Anderson and Natland¹ favour the idea that all plumes originate from the core–mantle boundary. But must they? The elegant experiments of Davaille *et al.*¹⁰ show that imagining plumes as narrow conduits from the core–mantle boundary might be a narrow view of convective patterns in the mantle.

B. Bourdon¹, N. M. Ribe², A. Stracke¹, A. Saal³ & S. Turner⁴

¹Institute of Isotope Geochemistry and Mineral Resources, Department of Earth Sciences, ETH Zürich, CH-8092 Zürich, Switzerland.

e-mail: bourdon@erdw.ethz.ch

²Laboratoire de Dynamique des Systèmes Géologiques, Institut de Physique du Globe de Paris-CNRS, 4, Place Jussieu, 75252 Paris cedex 05, France.

³Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA.

⁴GEMOC, Department of Earth and Planetary Sciences, Macquarie University, Sydney, New South Wales 2109, Australia.

1. Anderson, D. L. & Natland, J. H. Evidence for mantle plumes? *Nature* **450**, doi:10.1038/nature06376 (2007).
2. Bourdon, B., Ribe, N., Stracke, A., Saal, A. & Turner, S. P. Insights into the dynamics of mantle plumes from uranium-series geochemistry. *Nature* **444**, 713–717 (2006).
3. Campbell, I. H. & Kerr, A. C. The great plume debate: Testing the plume theory. *Chem. Geol.* **241**, 149–374 (2007).
4. Davies, G. F. A case for mantle plumes. *Chin. Sci. Bull.* **50**, 1541–1554 (2005).
5. Turcotte, D. L. & Phipps Morgan, J. in *Mantle Flow and Melt Generation at Mid-Ocean Ridges* (eds Phipps Morgan, J., Blackmann, D. K. & Sinton, J. M.) 155–182 (Geophysical Monograph 71, American Geophysical Union, 1992).
6. Bourdon, B., Turner, S. P. & Ribe, N. M. Partial melting and upwelling rates beneath the Azores from a U-series isotope perspective. *Earth Planet. Sci. Lett.* **239**, 42–56 (2005).
7. Sims, K. W. W. *et al.* Porosity of the melting zone and variations in the solid mantle upwelling rates beneath Hawaii: Inferences from ²³⁸U–²³⁰Th–²²⁶Ra, and ²³⁵U–²³¹Pa disequilibria. *Geochim. Cosmochim. Acta* **63**, 4119–4138 (1999).
8. Herzberg, C. *et al.* Temperatures in ambient mantle and plumes: Constraints from basalts, picrites, and komatiites. *Geochim. Geophys. Geosyst.* **8**, Q02006, doi:10.1029/2006GC001390 (2007).
9. Bourdon, B. & Sims, K. W. W. in *U-series Geochemistry* (eds Bourdon, B., Lundstrom, C., Henderson, G. & Turner, S. P.) 215–254 (Reviews in Mineralogy and Geochemistry Vol. 52, Geochemical Society and Mineralogical Society of America, Washington, 2003).
10. Davaille, A., Girard, F. & Le Bars, M. How to anchor hotspots in a convecting mantle? *Earth Planet. Sci. Lett.* **203**, 621–634 (2002).

doi:10.1038/nature06377

Producing primate embryonic stem cells by somatic cell nuclear transfer

J. A. Byrne^{1†}, D. A. Pedersen¹, L. L. Clepper¹, M. Nelson³, W. G. Sanger³, S. Gokhale³, D. P. Wolf¹ & S. M. Mitalipov^{1,2}

Derivation of embryonic stem (ES) cells genetically identical to a patient by somatic cell nuclear transfer (SCNT) holds the potential to cure or alleviate the symptoms of many degenerative diseases while circumventing concerns regarding rejection by the host immune system. However, the concept has only been achieved in the mouse, whereas inefficient reprogramming and poor embryonic development characterizes the results obtained in primates. Here, we used a modified SCNT approach to produce rhesus macaque blastocysts from adult skin fibroblasts, and successfully isolated two ES cell lines from these embryos. DNA analysis confirmed that nuclear DNA was identical to donor somatic cells and that mitochondrial DNA originated from oocytes. Both cell lines exhibited normal ES cell morphology, expressed key stem-cell markers, were transcriptionally similar to control ES cells and differentiated into multiple cell types *in vitro* and *in vivo*. Our results represent successful nuclear reprogramming of adult somatic cells into pluripotent ES cells and demonstrate proof-of-concept for therapeutic cloning in primates.

ES cells can differentiate into multiple cell types, representatives of which could be used in replacement therapy for ageing or diseased cells and tissues¹. However, ES cells derived from *in vitro* fertilized (IVF) embryos are genetically divergent from the patient (allogenic) and thus any resultant transplanted cell would be rejected without the continual application of immunosuppressive drugs². One way to resolve completely the transplant rejection issue would be to generate ES cells that are genetically identical to the patient. Recent reports suggest that somatic cells can be directly reprogrammed into ES-cell-like cells in mice, after transfection with key stem-cell-specific (stemness) genes^{3–5}. However, the ability to extrapolate this approach to other mammals, including primates, remains in question, and the propensity of such cells to develop into tumours owing to *c-myc* transgene reactivation remains a concern⁴. Alternatively, ES cells can be derived by epigenetic reprogramming of somatic cells via SCNT in spindle-free oocytes, a process commonly referred to as therapeutic cloning (Supplementary Fig. 1). However, despite encouraging results in the murine model^{6,7}, the feasibility of therapeutic cloning in primates remains in question. Although many mammals have been successfully produced by means of SCNT⁸ no successful primate reproductive cloning has been achieved^{9,10}. Moreover, the efficiency of blastocyst formation after human¹¹ and non-human primate⁹ SCNT has typically been very low, suggesting lack or incomplete nuclear reprogramming with existing SCNT protocols. We recently reported incomplete nuclear remodelling, including nuclear envelope breakdown and premature chromosome condensation, after standard SCNT in rhesus macaques (*Macaca mulatta*) and correlated this observation with a decline in maturation promoting factor activity¹². SCNT protocols designed to prevent premature cytoplasm activation and maturation promoting factor decline resulted in robust nuclear envelope breakdown and premature chromosome condensation and in significantly increased blastocyst development, suggesting that maturation promoting factor activity is instrumental for efficient nuclear reprogramming¹². Here we used these SCNT protocols to demonstrate therapeutic

cloning in the rhesus macaque and provide a translational model for testing the feasibility, therapeutic effectiveness and long-term safety of therapeutic cloning as a concept.

Embryo development and ES cell derivation

The primary culture of skin fibroblasts, used as the source of nuclear donor cells for SCNT, was established from a nine-year-old adult rhesus macaque male (male 1) housed at the Oregon National Primate Research Center (Supplementary Fig. 2). Mature metaphase II rhesus macaque oocytes were rendered spindle-free with the Oosight spindle imaging system that uses polarized light to visualize the oocyte meiotic spindle (Supplementary Fig. 3). Analysis of the removed karyoplasts, for the presence of the meiotic spindles, consistently confirmed a 100% efficiency of spindle removal using this approach. The donor fibroblast nuclei were introduced into cytoplasts by electrofusion, incubated for 2 h to allow nuclear remodelling to occur, and subsequently activated and cultured to the blastocyst stage as described previously¹² (see also Methods). We observed a 16% (35 out of 213) blastocyst formation rate with this nuclear donor cell line (Supplementary Table 1). SCNT blastocysts demonstrated a similar morphology to low-grade IVF blastocysts (Supplementary Fig. 4). Twenty expanded or hatching SCNT blastocysts were used for ES cell derivation via mechanical inner cell mass (ICM) isolation ($n = 2$), immunosurgery ($n = 15$) or direct culture of intact blastocysts ($n = 3$) on mouse embryonic fibroblast (MEF) feeder layers. Two ES cell lines (cloned rhesus embryonic stem: CRES-1 and CRES-2) were derived, both after immunosurgery (10% derivation efficiency from blastocysts). Overall, 304 oocytes collected from 14 rhesus macaque females were used to generate two ES cell lines, a 0.7% derivation efficiency from oocytes.

Genetic analysis and pluripotency

As the number of mitochondria, each with 16.6 kilobases (kb) of mitochondrial DNA¹³ (mtDNA), in the cytoplasm dwarfs any mitochondrial contribution from the donor somatic cells, embryos

¹Oregon National Primate Research Center and ²Oregon Stem Cell Center, Oregon Health & Science University, 505 N.W. 185th Avenue, Beaverton, Oregon 97006, USA. ³Munroe-Meyer Institute, 985450 Nebraska Medical Center, Omaha, Nebraska 68198, USA. ⁴Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. †Present address: Stanford Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Palo Alto, California 94304, USA.

derived by SCNT should predominantly, if not exclusively, possess mitochondria inherited from the oocyte¹⁴. Therefore, ES cells derived from SCNT embryos should contain mtDNA identical to the female providing the recipient cytoplasts and nuclear DNA genetically identical to the male providing the nuclear donor cells. To investigate whether the CRES-1 and CRES-2 cell lines contained the same nuclear DNA as the donor (male 1) fibroblasts, we performed microsatellite typing using 39 short tandem repeat (STR) loci¹⁵ and analysis of 56 single nucleotide polymorphisms (SNPs)¹⁶, 30 of which were informative for inheritance. Both the STR analysis—which included 25 common STR loci (Table 1) and 14 major histocompatibility complex (MHC)-linked STRs (Supplementary Table 2)—and the SNP analysis (Supplementary Table 3) demonstrated a complete match of both CRES lines to each other and to the nuclear DNA isolated from skin fibroblasts and peripheral blood leucocytes of male 1. In contrast, DNA obtained from the oocyte donor females for CRES-1 (female 1) and CRES-2 (female 2) demonstrated no significant similarity to CRES-1 or CRES-2 (Table 1 and Supplementary Tables 2 and 3). The genomic constitution of an IVF-derived rhesus macaque ES cell line (ORMES-22; ref. 17), and the ORMES-22 oocyte donor female (female 3) and sperm donor male (male 2), were also included to demonstrate STR allele inheritance (Table 1 and Supplementary Table 2).

To investigate whether the CRES-1 and CRES-2 cell lines contained the same mtDNA as their respective oocyte donor females, we performed mtDNA sequence analysis investigating an informative domain 1 (ID1) in the rhesus macaque mitochondrial D-loop hypervariable region 2 (RhDHV2). This RhDHV2 sequence contained multiple informative SNPs including at ID1 nucleotide positions 4, 22 and 28 (Fig. 1). Analysis of SNP22 (an A-to-G polymorphism) demonstrated that CRES-1 mtDNA was derived from the oocyte donor female 1 and not from the nuclear donor for CRES-1. Similarly, analysis of SNP4 (a C-to-T polymorphism) and SNP28 (an A-to-G polymorphism) confirmed that the CRES-2 mtDNA was derived from the CRES-2 oocyte donor female 2 and not from the nuclear donor. Thus, microsatellite, SNP and mtDNA analyses verified that CRES-1 and CRES-2 contained nuclear DNA genetically

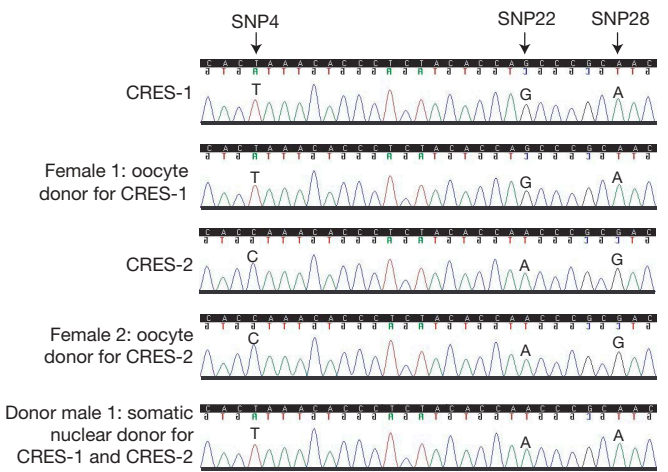


Figure 1 | Chromatograms of the rhesus macaque mitochondrial D-loop hypervariable region 2 informative domain 1 (RhDHV2 ID1). The ID1 sequence encompassed *Macaca mulatta* mtDNA nucleotide positions 451–480 (GenBank accession number NC_005943). Sequence analysis revealed that SNP22 (A-to-G) was informative for the mitochondrial inheritance for CRES-1 whereas SNP4 (C-to-T) and SNP28 (A-to-G) were informative for the mitochondrial inheritance for CRES-2. Results were confirmed by three independent sequences.

identical to the nuclear donor fibroblasts and mtDNA inherited from oocytes, a hallmark of SCNT-produced ES cells and offspring.

Both CRES lines demonstrated typical ES cell morphology (Fig. 2a–d), maintained an undifferentiated morphology after repeated manual passaging (>20 passages per line so far) and expressed key primate stemness markers including OCT4 (also called POU5F1), SSEA-4, TRA1-60 and TRA1-81 (Fig. 2). Moreover, transcripts of other stemness genes including *NANOG*, *SOX2*, *LEFTY1* (also called *LEFTY2*), *TDGF* and *TERT* were detected by polymerase chain reaction with reverse transcription (RT-PCR) analysis in both IVF-derived ES cell controls (ORMES-10 and ORMES-22) and CRES cell lines (Supplementary Fig. 5). Conventional cytogenetic G-banding

Table 1 | Short tandem repeat (STR) analysis of CRES cell lines

STR loci	Female 1 oocyte donor for CRES-1	Female 2 oocyte donor for CRES-2	Male 1 somatic nuclear donor for CRES-1 and CRES-2	CRES-1	CRES-2	Female 3 oocyte donor for ORMES-22	Male 2 sperm donor for ORMES-22	ORMES-22
Sex (AME)	XX	XX	XY	XY	XY	XX	XY	XX
D1S548	190/206	190/198	190/190	190/190	190/190	190/190	190/190	190/190
D2S1333	301/301	293/301	289/301	289/301	289/301	273/293	285/289	273/285
D3S1768	221/221	205/213	193/217	193/217	193/217	205/213	205/205	205/205
D4S2365	283/283	275/287	283/283	283/283	283/283	283/283	283/283	283/283
D4S413	131/131	133/145	131/139	131/139	131/139	131/145	125/141	131/141
D5S1457	136/136	132/136	132/136	132/136	132/136	132/136	132/140	136/140
D6S501	176/180	176/180	176/180	176/180	176/180	188/192	180/180	180/188
D7S513	191/205	205/209	189/191	189/191	189/191	189/217	193/199	199/217
D7S794	108/124	124/128	128/128	128/128	128/128	108/108	108/128	108/128
D8S1106	144/144	148/160	144/148	144/148	144/148	148/168	160/168	168/168
D9S921	183/195	183/191	179/179	179/179	179/179	183/195	175/195	183/195
D10S1412	157/166	160/160	157/157	157/157	157/157	157/157	160/160	157/160
D11S2002	256/256	256/256	260/264	260/264	260/264	252/252	256/260	252/256
D11S925	308/338	310/316	308/310	308/310	308/310	308/308	338/338	308/338
D12S364	282/290	282/288	281/290	281/290	281/290	282/290	268/296	268/290
D12S67	121/129	192/204	117/125	117/125	117/125	117/133	109/117	109/133
D13S765	228/240	212/220	228/256	228/256	228/256	216/236	228/228	228/236
D15S823	333/349	329/353	329/361	329/361	329/361	357/385	345/353	345/385
D16S403	164/168	156/158	158/164	158/164	158/164	152/164	152/152	152/164
D17S1300	232/280	244/280	272/276	272/276	272/276	248/252	228/284	252/284
D18S537	178/178	178/178	174/178	174/178	174/178	174/178	162/174	162/178
D18S72	306/308	306/322	306/308	306/308	306/308	308/308	306/308	308/308
D22S685	315/319	291/303	315/327	315/327	315/327	311/311	327/327	311/327
MFGT21	113/115	117/119	115/115	115/115	115/115	111/113	115/125	113/125
MFGT22	104/104	104/104	100/104	100/104	100/104	100/104	104/110	104/104

CRES-1 is an ES cell line isolated from a SCNT blastocyst produced by fusion of a cytoplast donated by female 1 and a skin fibroblast originated from male 1. Similarly, CRES-2 was derived from a cytoplast donated by female 2 and a donor somatic nucleus from male 1. ORMES-22 is an ES cell line derived from an embryo produced by IVF of a female 3 oocyte with a male 2 sperm. AME, amelogenin.

analysis of the nuclear donor fibroblasts used for SCNT (Supplementary Fig. 6) and the CRES-2 cell line (Fig. 3a) demonstrated a normal male rhesus macaque chromosome (42, XY) complement in all cells analysed. However, analysis of the CRES-1 cell line indicated the presence of three metaphase cells representing a hypodiploid clone characterized by loss of the Y chromosome (Supplementary Fig. 7), and seventeen cells representing a diploid clone characterized by an isochromosome comprised of two copies of the long arm of the Y chromosome (41,X[3]/42,X,i(Y)q10[17]) (Fig. 3b). Subsequent fluorescent *in situ* hybridization (FISH) analysis confirmed the G-banding findings; metaphase cells revealed the presence of a signal for bacterial artificial chromosome (BAC) CH250-283K14 on both arms of the Y chromosome (Fig. 3d), indicating the presence of the i(Y)q10 observed in the G-banding study. Additional studies were positive for loss of the Y chromosome in 12% of the CRES-1 cells analysed.

Transcriptional profiling

Global transcription profiles of three biological replicates each of male 1 skin fibroblasts (nuclear donor for both CRES lines), both CRES-1 and CRES-2 cell lines and two control ES cell lines derived from fertilized embryos (ORMES-10 and ORMES-22¹⁷) were examined by Affymetrix microarray analysis. For the primary microarray comparison (Supplementary Data 3), three types of analyses were performed: (1) replicates of each cell line were compared against each other, (2) each cell line was compared against the somatic donor cell line; and (3) each cell line was compared to a control IVF-derived ES

cell line (see Methods). For each comparison, the detected signal for each present 'P' probe set ($P < 0.05$) was plotted in a scatter graph, the number of present probe sets (PP) used was recorded and the correlation value was calculated. All comparisons of control ORMES biological replicates with each other demonstrated a correlation value of greater than 60% and all unrelated sample comparisons (that is, between ES cell and somatic cell biological replicates) demonstrated a correlation value of significantly less than 60%; therefore a correlation value of 60% or greater was considered indicative of a significant transcriptional correlation. When the replicates of the somatic donor cells were compared, 99% transcriptional correlation was observed (Supplementary Fig. 8a), suggesting that minimal artificial variation was introduced via the protocols used. Although it was not possible to determine with certainty the degree of technical versus biological variation between replicates, it should be noted that all samples were processed identically and at the same time, and the level of technical variation between the donor somatic cell samples was 1% or less, suggesting that most of the 20–30% transcriptional variation observed between ES cell replicates was biological in origin. If so, ES cells show significant transcriptional plasticity not observed in somatic cells (Supplementary Fig. 8a). However, further transcriptional profiling will be required to confirm this observation. Comparisons of the CRES cell lines to the somatic donor cells and control IVF-derived ES cells demonstrated that both CRES lines had fully reprogrammed into an ES cell transcriptional state, with no significant transcriptional correlation between the CRES lines and the donor somatic cells (Supplementary Fig. 8b) but a significant

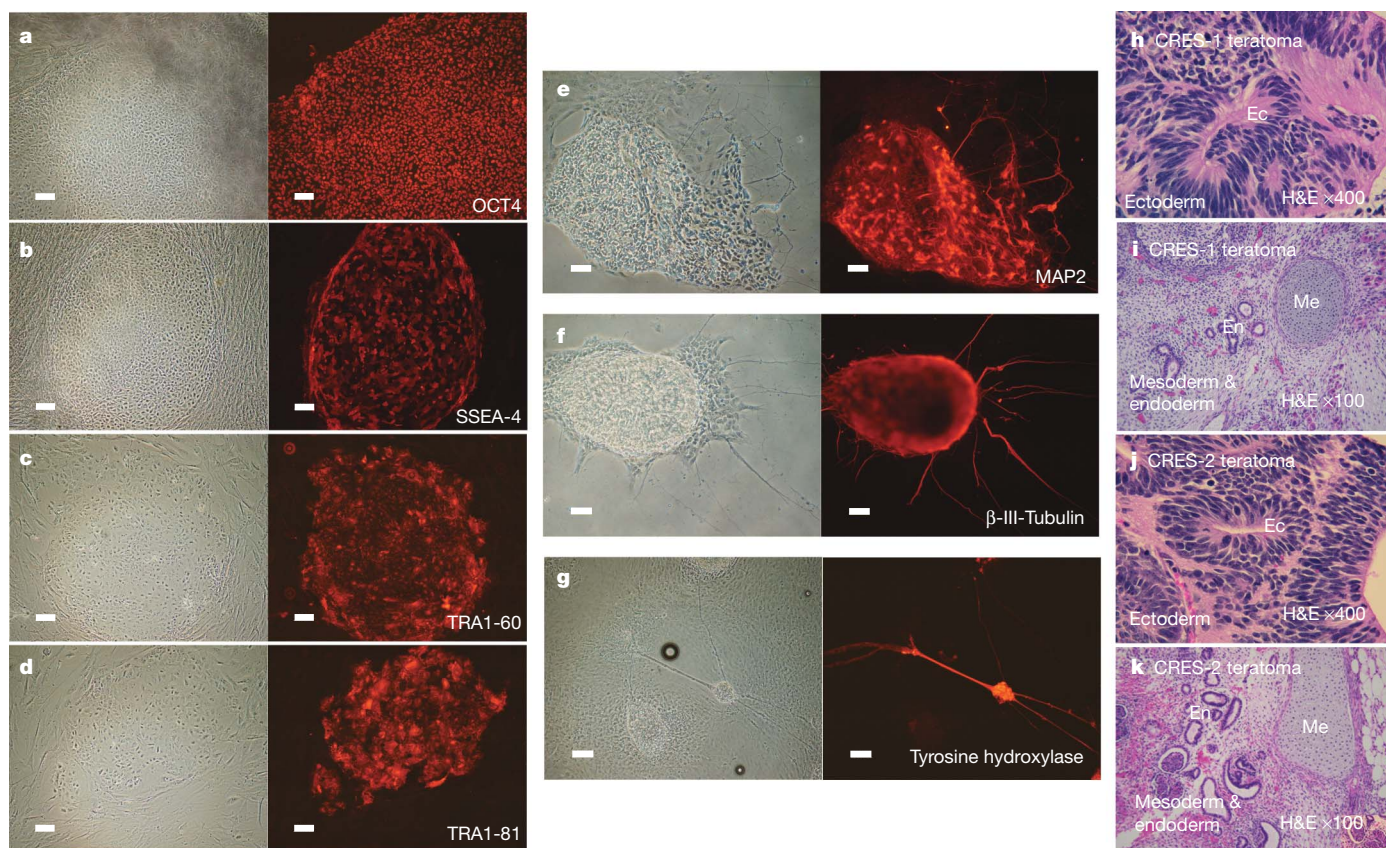


Figure 2 | Morphological and immunocytochemical analysis of CRES cells and their differentiated derivatives. **a–d**, Morphology of CRES colonies and expression of primate ES-cell-specific markers: OCT4, SSEA-4, TRA1-60 and TRA1-81. For each stemness marker analysed, a positive signal was only observed in CRES colonies, not in mouse embryonic fibroblast feeder cells, demonstrating antibody specificity. **e–g**, Cellular morphology and expression of neural markers in differentiated CRES cells, including microtubule-associated protein 2 (MAP2), β-III-tubulin and tyrosine

hydroxylase. The left hand column of **a–g** represents phase contrast images and the right hand column demonstrates marker expression as detected via immunocytochemistry. The scale bars in **a–d**, **g** represent 100 μm and the scale bars in **e**, **f** represent 50 μm. **h**, **j**, Presence of ectoderm (Ec)-derived neuroepithelium (×400, haematoxylin and eosin (H&E)) in teratomas produced by injection of CRES-1 and CRES-2 into SCID mice. **i**, **k**, Presence of mesoderm (Me)-derived cartilage and endoderm (En)-derived glandular epithelium (×100, H&E) in CRES-1 and CRES-2 teratomas.

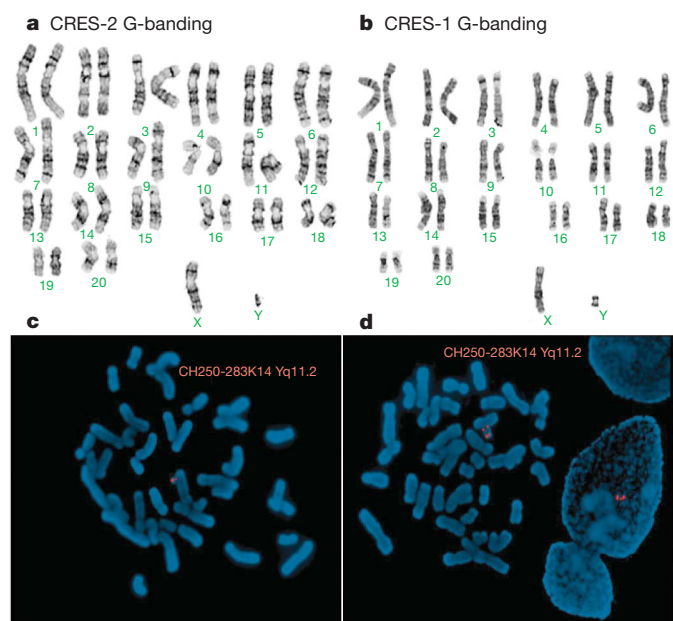


Figure 3 | Cytogenetic analysis of CRES cells. **a**, G-banding of CRES-2 cells demonstrating a normal euploid rhesus karyotype (42, XY). **b**, CRES-1 cell line with aneuploid karyotype, characterized by an isochromosome comprised of two copies of the long arm of the Y chromosome (41,X[3]/42,X,i(Y)(q10)[17]). **c**, FISH analysis of the nuclear donor fibroblasts demonstrating a normal karyotype with two red fluorescent signals on the long (q) arm of the Y chromosome. **d**, FISH analysis of CRES-1 cells indicating the presence of four signals for the Y chromosome long (q) arm, confirming the presence of the i(Y)(q10) aneuploidy.

correlation between CRES cells and the control ES cells (Supplementary Fig. 8c).

To identify ES-cell-specific genes, comparison analysis was performed between each of the three control ORMES-10 replicates and each of the three somatic donor cell replicates, to give a total of nine ES cell–somatic comparisons (Supplementary Data 4). This set of ES cell comparisons identified 4,998 somatic-cell-specific probe sets/genes (Supplementary Data 5) and 6,178 ES-cell-specific probe sets/genes (Supplementary Data 6 and Methods). Over 90% of the somatic-cell-specific genes were significantly downregulated in the CRES cell line replicates (Supplementary Table 4), and over 85% of the ES-cell-specific genes demonstrated significantly greater expression in the CRES cell line replicates (Table 2) when compared with the somatic donor cells. Transcriptional analysis of the control ORMES-22 replicates also demonstrated that over 90% of the somatic-specific genes had significantly less expression (Supplementary Table 4) and over

85% of the ES-cell-specific genes had significantly greater expression (Table 2) when compared with the somatic donor cells.

The final microarray analysis involved examining the level of expression of 12 putative rhesus macaque stemness genes identified in previous transcriptional profiling¹⁸. These putative stemness genes had the highest average fold change in gene expression when undifferentiated ES cell biological replicates were compared to their *in vitro* differentiated counterparts, and all 12 were significantly upregulated in the five different ES cell lines examined¹⁸. All 12 stemness genes were significantly upregulated in all of the ORMES-10, ORMES-22, CRES-1 and CRES-2 replicates (Supplementary Data 7), and the average fold change in gene expression for both CRES-1 and CRES-2 was comparable to that for ORMES-10 and ORMES-22 when compared to somatic donor cell replicates (Table 3). As a control, the relative fold change in expression for these putative stemness genes between the donor somatic cell replicates was insignificant (Table 3).

Our overall conclusion after analysis of global transcriptional profiles is that both CRES-1 and CRES-2 cells are transcriptionally similar to control ES cell lines derived from IVF-produced blastocysts.

Differentiation potential

To define pluripotency further, both CRES lines were exposed to conditions for cardiomyocyte differentiation *in vitro*¹⁷. CRES-1 and CRES-2 efficiently produced contracting aggregates (Supplementary Video) expressing markers of cardiac muscle tissue (Supplementary Fig. 9). Directed neural differentiation resulted in efficient formation of various neural phenotypes, demonstrating elongated cellular morphology and expression of neural markers, including microtubule-associated protein 2 (MAP2; Fig. 2e), β -III-tubulin (Fig. 2f) and tyrosine hydroxylase (Fig. 2g). When injected into severe combined immune deficiency (SCID) mice, both CRES lines formed teratomas and subsequent histological analysis identified representatives of all three germ layers (Fig. 2h–k), confirming their pluripotent status.

Discussion

Our results demonstrate successful reprogramming of primate somatic cells into pluripotent ES cells, provided that an efficient SCNT method is available. We achieved a significant increase in the blastocyst formation rate (from 1% to 16%) when the use of Hoechst 33342 was avoided during the oocyte spindle removal step. The detrimental effect of fluorochrome bisbenzimidazole (Hoechst 33342) and ultraviolet light on cytoplasm quality has been previously documented^{12,19}, and we speculate that the impaired blastocyst formation rate after conventional SCNT in primates may result from one or more of the following factors: Hoechst 33342 and/or ultraviolet damage to the relatively transparent primate oocyte; Hoechst 33342/UV-induced oocyte activation and/or maturation promoting factor degradation; reaction of the residual Hoechst 33342 in cytoplasm with the

Table 2 | Analysis of ES-cell-specific gene expression in rhesus macaque stem cell lines

Cell line	Biological replicate	Number of ES cell genes* upregulated compared to donor line replicate A†	Number of ES cell genes* upregulated compared to donor line replicate B†	Number of ES cell genes* upregulated compared to donor line replicate C†
Nuclear donor cells	A	N/A	21 (0.3%)	47 (0.8%)
Nuclear donor cells	B	30 (0.5%)	N/A	77 (1.2%)
Nuclear donor cells	C	18 (0.3%)	13 (0.2%)	N/A
ORMES-22	A	5,482 (89%)	5,388 (87%)	5,389 (87%)
ORMES-22	B	5,558 (90%)	5,607 (91%)	5,644 (91%)
ORMES-22	C	5,766 (93%)	5,672 (92%)	5,723 (93%)
CRES-1	A	5,974 (97%)	6,001 (97%)	5,984 (97%)
CRES-1	B	5,896 (95%)	5,919 (96%)	5,926 (96%)
CRES-1	C	5,748 (93%)	5,845 (95%)	5,784 (94%)
CRES-2	A	5,931 (96%)	5,843 (95%)	5,850 (95%)
CRES-2	B	5,658 (92%)	5,552 (90%)	5,483 (89%)
CRES-2	C	5,863 (95%)	5,933 (96%)	5,889 (95%)

* The general approximation when working with large numbers of probe sets is to assume that each probe set represents hybridization to a single gene. However, multiple probe sets can exist for certain genes, so the actual number of genes included in the analysis is significantly lower than the number of probe sets analysed.

† The percentage value represents the proportion of probe sets, out of the 6,178 ES-cell-specific probe sets, significantly upregulated after comparison analysis.

Table 3 | Expression analysis of putative rhesus macaque stemness genes

Affymetrix probe set ID	Stemness gene*	Nuclear donor cells fold change†	ORMES-10 fold change†	ORMES-22 fold change†	CRES-1 fold change†	CRES-2 fold change†
MmugDNA.26523.1.S1_s_at	<i>NFE2L3</i>	1	429	389	532	378
MmuSTS.2285.1.S1_at	<i>POU5F1 (OCT4)</i>	0	315	288	320	281
MmugDNA.9427.1.S1_at	<i>NR5A2</i>	1	282	310	278	325
MmugDNA.32128.1.S1_at	<i>NANOG</i>	1	246	180	256	190
MmuSTS.1436.1.S1_at	<i>LCK</i>	1	179	206	218	94
MmugDNA.11728.1.S1_at	<i>VTCN1</i>	1	245	139	153	125
MmugDNA.42677.1.S1_at	<i>DPPA4</i>	4	154	117	178	78
MmugDNA.28461.1.S1_at	<i>SLC12A1</i>	1	71	128	185	169
MmugDNA.6836.1.S1_at	<i>C14orf115 (LOS699513)</i>	1	81	87	85	64
MmuSTS.3122.1.S1_at	<i>MYRIP</i>	0	71	51	75	53
MmugDNA.15193.1.S1_at	<i>ADH4</i>	0	68	13	52	58
MmuSTS.2310.1.S1_at	<i>PRDM14</i>	1	28	44	46	40

* The stemness genes were identified in previous transcriptional profiling of rhesus macaque ES cell lines¹⁸. These 12 genes had the highest average fold change in gene expression of five different rhesus macaque ES cell lines.

† Microarray suite-5 (MAS-5) statistical analysis was performed to calculate the signal log ratio (SLR) for each cell line in comparison to the somatic donor cell line biological replicates, and the averaged gene expression fold change between compared samples was calculated from the SLR using the formula: fold change = (2^{SLR}) . All 12 stemness genes were significantly upregulated ($P < 0.002$) in CRES cell lines.

introduced donor cell DNA, thereby impairing reprogramming; and/or Hoechst 33342 contact with mitochondrial DNA, thus reducing cytoplasm mitochondrial function.

Recognizing the importance of high-quality cytoplasts for successful reprogramming, we have sought non-invasive approaches for spindle detection and removal. 'Blind' enucleation techniques involving 'squish'²⁰ or 'one-step manipulation'²¹ were inefficient, at least in our hands, because they failed to enucleate all oocytes. In fact, our initial effort to derive ES cells from SCNT blastocysts produced with these protocols resulted in the isolation of parthenogenetic ES cells owing to failed spindle removal. Fortunately, recent developments in high-performance imaging resulted in an Oosight spindle imaging system supporting rapid and highly efficient real-time enucleation of primate oocytes. Introduction of the donor nucleus can be accomplished by either direct injection or electrofusion—the latter was dictated here by the relatively large donor cells used. The SCNT blastocyst development rate was significantly lower than fertilized controls but within the 7–29% range described by us when fetal fibroblasts, adult ear fibroblasts, cumulus and oviductal epithelial cells were used as nuclear donor cells in the monkey¹². Blastocysts produced by SCNT usually demonstrated poor morphology and have, so far, failed to support a term pregnancy after embryo transfer despite considerable efforts^{21,22} (D.P.W., unpublished data). This result may not be predictive for ES cell isolation, however, as the requirements for reproductive and therapeutic cloning are different in that a normal trophectoderm is not required for the latter²³.

With an adequate supply of SCNT blastocysts, the final challenge in therapeutic cloning is ES cell isolation, and although several methods were examined, the conventional method involving immunosurgical dispersal of the trophectoderm seemed to be the best. The derivation efficiency in the present study is within the range reported in the mouse, where the ES cell derivation efficiency from SCNT embryos was 0.2–3.4% per oocyte and 4–10% per blastocyst^{6,24–27}.

Regarding the origin of the CRES lines, in addition to the 100% spindle removal efficiency, karyotype, microsatellite and SNP analyses confirmed that both CRES lines originated from SCNT embryos and not from parthenotes. CRES lines demonstrated typical ES cell morphology, self-renewal capacity and expression of stemness markers. These cell lines were also transcriptionally similar to ES cells derived from fertilized blastocysts, and pluripotent, as demonstrated by the generation of representatives of all three germ layers after *in vivo* teratoma formation. Our results parallel findings in the mouse^{28,29} and confirm the possibility of complete nuclear reprogramming of somatic cells into pluripotent ES cells via SCNT. The CRES-1 cell line revealed a translocation of unknown origin characterized by an isochromosome comprised of two copies of the long arm of chromosome Y, whereas CRES-2 exhibited a normal euploid male karyotype. In our previous study²¹, five out of seven analysed SCNT blastocysts were karyotypically normal with two embryos

demonstrating aneuploidy. Karyotyping individual SCNT embryos based on either biopsied blastomere or trophectodermal cell analysis followed by ES cell isolation could address whether chromosomal abnormalities in SCNT-derived ES cells originate from aneuploid embryos or occur during ES cell isolation and culture.

The primary rationale for therapeutic cloning is transplantation of histocompatible ES-cell-derived phenotypes back into the patient. The MHC profile of both CRES lines perfectly matched the donor male in all MHC loci examined via microsatellite analysis, suggesting that transplantation of differentiated derivatives back into the donor animal would not lead to rejection. It is possible that immune response will occur in the donor animal resulting from the epitopes derived from the allogenic recipient oocyte mitochondria contribution; although preliminary research with cloned bovine and porcine cells and tissues with allogenic mtDNA suggests that grafting back into the nuclear donor organism can occur without destruction by the immune system^{30,31}. The use of biological materials from other species, such as fetal bovine serum, mouse embryonic fibroblasts and guinea-pig complement, may also contaminate ES cells with xeno-epitopes and pathogens that would make such cells unsuitable for clinical applications. Other possibly minor factors that should nevertheless also be investigated include the problems that may arise from the lack of sperm-derived centrosomes, the possible effects of non-random X-inactivation and the possibility that cloned ES cells may possess the shortened telomeres of their donor somatic cell source. The possibility has also been suggested that SCNT embryos are epigenetically abnormal based on DNA methylation patterns³². Therefore, the epigenetic status of CRES cell lines including imprinted gene expression would also be an interesting area for future research.

Assuming that our modified SCNT protocols are applicable to humans, the low efficiency of ES cell derivation (0.7%) along with restricted availability and high cost of human oocytes suggest that a significant increase in the overall SCNT embryo generation and ES cell derivation rates would be required before a clinical implementation of human therapeutic cloning could occur. Recently, mouse ES cells were relatively efficiently cloned using aged, failed-to-fertilize oocytes³³, and substantial numbers of such human oocytes augmented by immature oocytes are routinely discarded, representing an untested source for human SCNT. In addition to therapeutic cloning, human SCNT-derived ES cells would also be extremely valuable in studies devoted to understanding disease mechanisms and developing possible treatments through *in vitro* experimentation³⁴.

METHODS SUMMARY

A primary culture of fibroblasts was established from a skin biopsy of an adult rhesus macaque male (male 1) and prepared for SCNT as previously described⁹. Mature metaphase II oocytes were rendered spindle-free using the Oosight imaging system (CRI, Inc.) and a donor somatic cell nucleus was introduced

into a cytoplasm through electrofusion. Reconstructed embryos were activated 2 h after fusion by exposure to 5 μ M ionomycin (CalBiochem) for 5 min followed by a 5-h incubation in 2 mM 6-dimethylaminopurine (DMAP), placed in HECM-9 medium and cultured at 37 °C in 6% CO₂, 5% O₂ and 89% N₂ until the expanded blastocyst stage. The ICMs of selected SCNT blastocysts were plated onto MEF feeder layers and cultured in ES cell culture medium for 5–7 days. ICMs that attached to the feeder layer and initiated outgrowth were manually dissociated into small clumps with a microscalpel and replated onto fresh MEFs. After the first passage, colonies with ES-cell-like morphology were selected for further propagation, characterization, low-temperature storage and *in vitro* and *in vivo* differentiation, as previously described¹⁷.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 May; accepted 9 October 2007.

- McKay, R. Stem cells—hype and hope. *Nature* **406**, 361–364 (2000).
- Drukker, M. & Benvenisty, N. The immunogenicity of human embryonic stem-derived cells. *Trends Biotechnol.* **22**, 136–141 (2004).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006); published online 10 August 2006.
- Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317 (2007); published online 6 June 2007.
- Wernig, M. *et al.* *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–324 (2007); published online 6 June 2007.
- Rideout, W. M. III, Hochedlinger, K., Kyba, M., Daley, G. Q. & Jaenisch, R. Correction of a genetic defect by nuclear transplantation and combined cell and gene therapy. *Cell* **109**, 17–27 (2002).
- Barberi, T. *et al.* Neural subtype specification of fertilization and nuclear transfer embryonic stem cells and application in parkinsonian mice. *Nature Biotechnol.* **21**, 1200–1207 (2003).
- Wilmut, I. *et al.* Somatic cell nuclear transfer. *Nature* **419**, 583–586 (2002).
- Mitalipov, S. M., Yeoman, R. R., Nusser, K. D. & Wolf, D. P. Rhesus monkey embryos produced by nuclear transfer from embryonic blastomeres or somatic cells. *Biol. Reprod.* **66**, 1367–1373 (2002).
- Simerly, C. *et al.* Molecular correlates of primate nuclear transfer failures. *Science* **300**, 297 (2003).
- Hall, V. J. *et al.* Developmental competence of human *in vitro* aged oocytes as host cells for nuclear transfer. *Hum. Reprod.* **22**, 52–62 (2007).
- Mitalipov, S. M. *et al.* Reprogramming following somatic cell nuclear transfer in primates is dependent upon nuclear remodeling. *Hum. Reprod.* **22**, 2232–2242 (2007).
- Birky, C. W. Jr. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc. Natl Acad. Sci. USA* **92**, 11331–11338 (1995).
- Bowles, E. J., Campbell, K. H. & St John, J. C. Nuclear transfer: preservation of a nuclear genome at the expense of its associated mtDNA genome(s). *Curr. Top. Dev. Biol.* **77**, 251–290 (2007).
- Penedo, M. C. *et al.* Microsatellite typing of the rhesus macaque MHC region. *Immunogenetics* **57**, 198–209 (2005).
- Ferguson, B. *et al.* Single nucleotide polymorphisms (SNPs) distinguish Indian-origin and Chinese-origin rhesus macaques (*Macaca mulatta*). *BMC Genomics* **8**, 43 (2007).
- Mitalipov, S. *et al.* Isolation and characterization of novel rhesus monkey embryonic stem cell lines. *Stem Cells* **24**, 2177–2186 (2006).
- Byrne, J. A., Mitalipov, S. M., Clepper, L. & Wolf, D. P. Transcriptional profiling of rhesus monkey embryonic stem cells. *Biol. Reprod.* **75**, 908–915 (2006).
- Byrne, J. A., Clepper, L., Wolf, D. P. & Mitalipov, S. M. in *International Society for Stem Cell Research, 5th Annual meeting* 52 (Cairns, Queensland, Australia, 2007).
- Simerly, C. *et al.* Embryogenesis and blastocyst development after somatic cell nuclear transfer in nonhuman primates: overcoming defects caused by meiotic spindle extraction. *Dev. Biol.* **276**, 237–252 (2004).
- Zhou, Q. *et al.* A comparative approach to somatic cell nuclear transfer in the rhesus monkey. *Hum. Reprod.* **21**, 2564–2571 (2006).
- Ng, S. C. *et al.* The first cell cycle after transfer of somatic cell nuclei in a non-human primate. *Development* **131**, 2475–2484 (2004).
- Yang, X. *et al.* Nuclear reprogramming of cloned embryos and its implications for therapeutic cloning. *Nature Genet.* **39**, 295–302 (2007).
- Munsie, M. J. *et al.* Isolation of pluripotent embryonic stem cells from reprogrammed adult mouse somatic cell nuclei. *Curr. Biol.* **10**, 989–992 (2000).
- Wakayama, T. *et al.* Differentiation of embryonic stem cell lines generated from adult somatic cells by nuclear transfer. *Science* **292**, 740–743 (2001).
- Hochedlinger, K. & Jaenisch, R. Monoclonal mice generated by nuclear transfer from mature B and T donor cells. *Nature* **415**, 1035–1038 (2002).
- Mombaerts, P. Therapeutic cloning in the mouse. *Proc. Natl Acad. Sci. USA* **100**, 11924–11925 (2003); published online 29 August 2003.
- Wakayama, S. *et al.* Equivalency of nuclear transfer-derived embryonic stem cells to those derived from fertilized mouse blastocysts. *Stem Cells* **24**, 2023–2033 (2006); published online 11 May 2006.
- Brambrink, T., Hochedlinger, K., Bell, G. & Jaenisch, R. ES cells derived from cloned and fertilized blastocysts are transcriptionally and functionally indistinguishable. *Proc. Natl Acad. Sci. USA* **103**, 933–938 (2006); published online 17 January 2006.
- Lanza, R. P. *et al.* Generation of histocompatible tissues using nuclear transplantation. *Nature Biotechnol.* **20**, 689–696 (2002); published online 3 June 2002.
- Martin, M. J. *et al.* Skin graft survival in genetically identical cloned pigs. *Cloning Stem Cells* **5**, 117–121 (2003).
- Yang, J. *et al.* Epigenetic marks in cloned rhesus monkey embryos: comparison with counterparts produced *in vitro*. *Biol. Reprod.* **76**, 36–42 (2007).
- Wakayama, S. *et al.* Establishment of mouse embryonic stem cell lines from somatic cell nuclei by nuclear transfer into aged, fertilization-failure mouse oocytes. *Curr. Biol.* **17**, R120–R121 (2007).
- Byrne, J., Mitalipov, S. & Wolf, D. Current progress with primate embryonic stem cells. *Curr. Stem Cell Res. Ther.* **1**, 127–138 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors acknowledge the Division of Animal Resources and the Endocrine Services Cores at the Oregon National Primate Research Center for assistance and technical services. We thank M. Sparman, C. Ramsey and V. Dighe of the Assisted Reproductive Technology Core for their embryological and logistical assistance; J. Fanton and D. Jacobs for laparoscopic oocyte retrievals; B. Ferguson for performing the SNP analysis; C. Penedo for microsatellite analysis; and R. Stouffer, M. Grompe and R. Reijo Pera for reviewing this manuscript. Microarray assays were performed in the Affymetrix Microarray Core of the OHSU Gene Microarray Shared Resource. This study was supported by funds from ONPRC and NIH grants to S. Mitalipov, R. Stouffer and D. Dorsa.

Author Contributions S.M.M. and J.A.B. designed experiments, conducted SCNT and ES cell derivation. L.L.C. performed DNA/RNA isolations and stemness gene expression. J.A.B. analysed the microarray data and performed the mitochondrial DNA analysis. D.A.P. assisted with ES cell derivation and performed ES cell culture, characterization and differentiation. W.G.S. and M.N. performed the cytogenetic analysis. S.G. analysed teratomas. S.M.M., J.A.B. and D.P.W. analysed the data and wrote the paper.

Author Information Microarray data, including CEL and CHP files, and Supplementary Data files containing microarray analyses (Supplementary Data 3–7) have been deposited in the Gene Expression Omnibus (GEO) database with accession number GSE7748 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7748>). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.M.M. (mitalipo@ohsu.edu).

METHODS

Animals. Adult rhesus macaques housed in individual cages were used in this study. All animal procedures were approved by the Institutional Animal Care and Use Committee at the ONPRC/OHSU.

Ovarian stimulation of rhesus macaques. Controlled ovarian stimulation and oocyte recovery was performed as has previously been described³⁵. Briefly, cycling females were subjected to follicular stimulation using twice-daily intramuscular injections of recombinant human FSH as well as concurrent treatment with Antide, a GnRH antagonist, for days 8–9. Unless indicated otherwise, all reagents were from Sigma-Aldrich and all hormones were from Ares Advanced Technologies Inc. Females received recombinant human luteinizing hormone on days 7–9 and recombinant human chorionic gonadotrophin (hCG) on day 10.

Somatic cell nuclear transfer. A primary culture of fibroblasts was established from an adult rhesus macaque male (male 1) as previously described⁹. Briefly, a small skin biopsy was surgically derived, washed in Ca^{2+} - and Mg^{2+} -free Dulbecco PBS (Invitrogen) and minced into pieces before incubation in Dulbecco Modified Eagle's Medium (DMEM, Invitrogen) containing 1 mg ml⁻¹ collagenase IV (Invitrogen) at 37 °C in 5% CO₂ for 40 min. Tissue pieces were then vortexed, washed, seeded into 75 cm³ cell culture flasks (Corning) containing DMEM supplemented with 100 IU ml⁻¹ penicillin, 100 µg ml⁻¹ streptomycin (Invitrogen), 10% FBS (DMEM/FBS culture media) and cultured at 37 °C in 5% CO₂ until reaching confluency. Fibroblasts were then disaggregated with trypsin treatment and frozen down in aliquots of 1×10^6 cells in medium containing 10% dimethyl sulfoxide (DMSO).

Fibroblasts were subsequently thawed, plated onto 4-well dishes (Nunc) and cultured under standard conditions until reaching 50–90% confluency. Cells were then synchronized in the G0/G1 phase of the cell cycle by culturing in DMEM medium with 0.5% FBS for 4 days before SCNT. Cumulus–oocyte complexes were collected from anaesthetized animals by laparoscopic follicular aspiration (28–29 h after hCG) and placed in TALP/HEPES medium³⁶ (modified Tyrode solution with albumin, lactate and pyruvate) containing 0.3% BSA (TH3) at 37 °C. Oocytes were stripped of cumulus cells by mechanical pipetting after brief exposure (<1 min) to hyaluronidase (0.5 mg ml⁻¹) and placed in chemically defined, protein-free HECM-9 medium (hamster embryo culture medium)³⁷ at 37 °C in 6% CO₂, 5% O₂ and 89% N₂ until further use.

Recipient MII oocytes were transferred to 30 µl manipulation droplets of TH3 with 5 µg ml⁻¹ cytochalasin B on a glass bottom manipulation dish (<http://www.willcowells.com>) covered with paraffin oil (Zander IVF) and incubated at 37 °C for 10–15 min before spindle removal. The chamber was then mounted on an inverted microscope (Olympus) equipped with the Oosight imaging system (CRI, Inc.), glass stage warmer (Tokai Hit, <http://www.tokaihit.com>) and Narishige micromanipulators. The spindle was located and extracted by aspiration into an enucleation pipette (20–25 µm outer diameter). Metaphase spindle removal was confirmed by its presence in the enucleation pipette. The Oosight imaging system allows non-invasive, polarized light imaging and detection of the spindle based on birefringence. Using this innovative approach, we were able to locate and quickly remove the oocyte spindle real-time with 100% efficiency. After oocyte spindle removal a disaggregated donor somatic cell was aspirated into a micropipette and placed into the perivitelline space of the cytoplasm on the side opposite the first polar body. Cell fusion was induced by two 50 µs DC pulses of 2.7 kV cm⁻¹ (Electro Square Porator T-820, BTX, Inc.) in 0.25 M D-sorbitol buffer containing 0.1 mM calcium acetate, 0.5 mM magnesium acetate, 0.5 mM HEPES and 1 mg ml⁻¹ fatty-acid-free BSA. Successful fusion was confirmed visually 30 min after electroporation by the disappearance of the donor cell in the perivitelline space. All nuclear transfer micromanipulation and fusion procedures were conducted on microscope stage warmers (Tokai Hit) maintaining 37 °C. Reconstructed embryos were activated by exposure to 5 µM ionomycin for 5 min in TALP/HEPES medium supplemented with 1 mg ml⁻¹ fatty-acid-free BSA and then transferred for 5 min in TALP/HEPES medium supplemented with 30 mg ml⁻¹ fatty-acid-free BSA and 2 mM 6-dimethylaminopurine (DMAP) followed by a 5 h incubation in HECM-9 medium containing 2 mM DMAP at 37 °C in 6% CO₂. Activated oocytes were placed in 4-well dishes containing HECM-9 medium supplemented with 10% FBS and 12 µM β-mercaptoethanol (BME) and cultured at 37 °C in 6% CO₂, 5% O₂ and 89% N₂ for a maximum of 10 days with medium change every other day.

ES cell derivation and culture. Zonae pellucidae of selected expanded SCNT blastocysts were removed by brief exposure (45–60 s) to 0.5% pronase in TH3 medium. A small proportion of embryos were either transferred directly to MEFs as whole blastocysts or after mechanical dissection of the ICM. Remaining blastocysts were subjected to immunosurgical isolation of the ICMs as previously described¹⁷. Briefly, zona-free blastocysts were exposed to rabbit anti-rhesus spleen serum (a gift from J. A. Thomson) for 30 min at 37 °C. After extensive washing in TH3, embryos were incubated in guinea-pig complement

reconstituted with HECM-9 (1:2, v/v) for an additional 30 min at 37 °C. Partially lysed trophoblast cells were mechanically dispersed by gentle pipetting with a small bore pipette (125 µm inner diameter; Stripper pipette, Midatlantic Diagnostics Inc.) followed by the rinsing of ICMs three times with TH3 medium. Isolated ICMs were plated onto Nunc 4-well dishes containing mitotically inactivated feeder layers consisting of MEFs and cultured in DMEM/F12 medium with glucose and without sodium pyruvate supplemented with 1% nonessential amino acids, 2 mM L-glutamine, 0.1 mM β-mercaptoethanol and 15% FBS at 37 °C, 3% CO₂, 5% O₂ and 92% N₂. ICMs that attached to the feeder layer and initiated outgrowth were manually dissociated into small cell clumps with a microscalpel and replated onto new MEFs. After the first passage, colonies with ES-cell-like morphology were selected for further propagation, characterization and low-temperature storage as previously described¹⁷. Medium was changed daily and ES cell colonies were split every 5–7 days by manual dissociation and replating collected clumps onto dishes with fresh MEFs.

In vitro and in vivo differentiation of ES cells. For embryoid body formation, entire ES cell colonies were loosely detached from feeder cells and transferred into feeder-free, 6-well, ultra-low adhesion plates (Corning Costar) and cultured in suspension in ES cell medium for 5–7 days. To induce further differentiation, embryoid bodies were transferred into collagen-coated, 6-well culture dishes (Becton Dickinson) to allow attachment. For neuronal differentiation, medium was replaced with serum-free DMEM/F12 containing ITS supplement (insulin, transferrin and sodium selenite, Invitrogen) and fibronectin (5 µg ml⁻¹; Invitrogen)³⁸. Cultures were maintained for 7 days, with medium replenishment every 2 days. The resulting cultures were disaggregated with collagenase or trypsin treatment and replated onto polyornithine- and laminin-coated plates or glass coverslips in N2 medium consisting of DMEM/F12 supplemented with laminin (1 µg ml⁻¹; Invitrogen), bFGF (10 ng ml⁻¹; R&D Systems), and N2 supplement (Invitrogen). Cultures were maintained for an additional 7 days with a daily medium change. After 7 days, bFGF was omitted from the medium and cultures were maintained for an additional 7–12 days to induce differentiation into mature neuronal phenotype. Differentiation into cardiac cells was initiated by embryoid body formation in suspension as described above. Embryoid bodies were then plated into collagen-coated dishes and cultures were maintained in ES cell medium for 2–4 weeks. In order to obtain teratomas, 3–5 million undifferentiated ES cells from each cell line were harvested and injected into the hind-leg muscle or subcutaneously into 4-week-old, SCID, beige male mice using an 18 gauge needle. Six to seven weeks after injection, mice were killed and teratomas were dissected, sectioned and histologically characterized for the presence of representative tissues of all three germ layers.

Immunocytochemical procedures. Undifferentiated and differentiated ES cells were fixed in 4% paraformaldehyde for 20 min. After permeabilization with 0.2% Triton X-100 and 0.1% Tween-20, nonspecific reactions were blocked with 10% normal serum (Jackson ImmunoResearch Laboratories, Inc.). Cells were then incubated for 40 min in primary antibodies, washed three times and exposed to secondary antibodies conjugated with fluorochromes (Jackson ImmunoResearch) before co-staining with 2 µg ml⁻¹ 4',6-diamidino-2-phenylindole (DAPI) for 10 min, whole-mounting onto slides and examination under epifluorescence microscopy. Primary antibodies for OCT4, SSEA-4, TRA1-60 and TRA1-81 were from Santa Cruz Biotechnology Inc. Neural-specific antibodies including microtubule-associated protein (MAP2C), β-III-tubulin and tyrosine hydroxylase were from Chemicon International Inc.

Cytogenetic analysis. Mitotically active ES cells in log phase were incubated with 120 ng ml⁻¹ ethidium bromide for 40 min at 37 °C, 5% CO₂, followed by 120 ng ml⁻¹ colcemid (Invitrogen) treatment for 20–40 min. Cells were then dislodged with 0.25% trypsin, and centrifuged at 200g for 8 min. The cell pellet was gently re-suspended in 0.075 M KCl solution and incubated for 20 min at 37 °C followed by fixation with methanol:glacial acetic acid (3:1) solution. Cytogenetic analysis was performed on 20 metaphase cells from each ES cell line following standard GTW-banding procedures. Images were acquired using the Cytovision Image Analysis System (Applied Imaging) and karyotypes were arranged as previously described³⁹. FISH analysis was performed on metaphase cells from the donor somatic cell line and CRES-1 using BAC CH250-283K14, specific for the rhesus macaque Yq11.21 region. The clone was obtained from the CHORI-250 rhesus macaque BAC library (Children's Hospital Oakland Research Institute, BACPAC Resources), grown in LB media supplemented with 25 µg ml⁻¹ chloramphenicol and isolated using a mini-prep protocol adapted from the standard Qiagen plasmid purification method (Qiagen Inc.). Probes were nick-translated with rhodamine-5-dUTP (Enzo Life Sciences) using a modification of the manufacturer's protocol (Vysis). Approximately 200 ng of the probe was precipitated along with 2 µg human Cot-I DNA (Invitrogen). Before hybridization, slides were pre-treated at 73 °C in 2× SSC for 2 min, followed by digestion at 37 °C in a protease solution for 5 min, post-fixation in 10% buffered formaldehyde with 2 M MgCl₂ and dehydration. The slide and

probe mixture were co-denatured at 74 °C for 2 min and incubated overnight at 37 °C using the HYBrite system (Vysis). Images were acquired using the Cytovision image analysis system.

RT-PCR. Total RNA was extracted from cells using an RNA purification kit (Invitrogen) according to the manufacturer's instructions. Total RNA was treated with DNase I before cDNA preparation using the SuperScript III first-strand synthesis system for RT-PCR (Invitrogen) according to the manufacturer's instructions. The first strand cDNA was further amplified by PCR using individual primer pairs for specific genes. The primer sequence, annealing temperature and amplicon size for each pair of primers are listed in Supplementary Table 5. All PCR samples were analysed by electrophoresis on 2% agarose gel containing 0.5 µg ml⁻¹ ethidium bromide and visualized on a transilluminator.

Microsatellite analysis. For STR genotyping, DNA was extracted from blood or cultured cells using commercial kits (Gentra). Six multiplexed PCR reactions were set up for the amplification of 39 markers representing 25 autosomal loci and 14 autosomal, MHC-linked loci. On the basis of the published rhesus macaque linkage map⁴⁰, these markers are distributed in 19 chromosomes. Two of the markers included in the panel, MFMT21 and MFMT22⁴¹, were developed from *Macaca fuscata* and do not have a chromosome assignment. PCRs were set up in 25 µl reactions containing 30–60 ng DNA, 2.5 mM MgCl₂, 200 µM dNTPs, 1× PCR buffer II, 0.5 U AmpliTaq (Applied Biosystems) and fluorescence-labelled primers in concentrations ranging from 0.06 to 0.9 µM, as required for each multiplex PCR. Cycling conditions consisted of four cycles of 1 min at 94 °C, 30 s at 58 °C, 30 s at 72 °C, followed by 25 cycles of 45 s at 94 °C, 30 s at 58 °C, 30 s at 72 °C and a final extension at 72 °C for 30 min. PCR products were separated by capillary electrophoresis on the ABI 3730 DNA analyser (Applied Biosystems) according to the manufacturer's instructions. Fragment size analysis and genotyping was done with the computer software STRand (<http://www.vgl.ucdavis.edu/informatics/STRand/>). Primer sequences for MHC-linked STRs 9P06, 246K06, 162B17(A and B), 151L13, 268P23 and 222I18 were designed from the corresponding rhesus macaque BAC clone sequences deposited in GenBank (accession numbers AC148662, AC148696, AC148683, AC148682, AC148698 and AC148689, respectively). Loci identified by the letter 'D' prefix were amplified using heterologous human primers.

SNP analysis. SNP analysis was performed as previously described¹⁶. Briefly, SNP genotyping was performed using iPLEX reagents and protocols for multiplex PCR, single base primer extension and generation of mass spectra, as per the manufacturer's instructions (for complete details see iPLEX Application Note, Sequenom). Four multiplexed assays containing 56 SNPs were included, 30 which were informative for CRES inheritance (see Supplementary Table 3). All 56 SNPs demonstrated a 100% match between male 1 donor cells and the CRES-1 and CRES-2 cell lines. Initial multiplexed PCRs were performed in 5 µl reactions on 384-well plates containing 5 ng of genomic DNA. Reactions contained 0.5 U HotStar Taq polymerase (QIAGEN), 100 nM primers, 1.25× HotStar Taq buffer, 1.625 mM MgCl₂ and 500 µM dNTPs. After enzyme activation at 94 °C for 15 min, DNA was amplified with 45 cycles of 94 °C × 20 s, 56 °C × 30 s, 72 °C × 1 min, followed by a 3-min extension at 72 °C. Unincorporated dNTPs were removed using shrimp alkaline phosphatase (0.3 U, Sequenom). Single-base extension was carried out by addition of single base primer extension primers at concentrations from 0.625 µM to 1.25 µM using iPLEX enzyme and buffers (Sequenom). Single base primer extension products were measured using the MassARRAY Compact system, and mass spectra were analysed using TYPER software (Sequenom).

Mitochondrial DNA analysis. DNA was extracted from cell lines using commercial kits (Gentra). The rhesus macaque mitochondrial D-loop hypervariable region 2 (RhDHV2) sequence was amplified for each sample using primers RhDF2 (5'-TAACATATCCGATCAGAGCC-3') and RhDR (5'-TTAAACACCTCTACGCCG-3')⁴². PCR for each sample was performed using Platinum PCR SuperMix (Invitrogen) containing 0.5 µM of each primer (final volume 50 µl). Reaction conditions were initial denaturation at 94 °C for 2 min; 35 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s, extension at 72 °C for 90 s and a final extension at 72 °C for 3 min, generating 544 bp of sequence covering the RhDHV2 region. Products from these reactions were then sequenced to determine polymorphic variation in the RhDHV2 region by direct sequencing. The informative domain 1 (ID1) sequence encompassing *Macaca mulatta* mtDNA nucleotide positions 451–480 (GenBank NC_005943) was identified as containing SNPs informative for the mitochondrial inheritance of both CRES-1 and CRES-2. Each ID1 sequence was confirmed by three other sequencing reactions and all of the RhDHV2 chromatograms used in this project were obtained with Sequencher v. 4.7 (GeneCodes).

Microarray analysis. The MIAME (Minimum Information About a Microarray Experiment) guidelines for microarray research⁴³ were incorporated into the design and implementation of these studies. All the microarray comparison analysis data files cited in the text are available online at the Gene

Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7748>). All the microarray comparison analysis data cited in the text are available at <ftp://ftp.ncbi.nih.gov/pub/geo/DATA/supplementary/series/GSE7748>. Total RNA was isolated from cell colonies selected for the appropriate ES cell morphology (flat monolayer colony with distinctive cobblestone cell morphology and a high nucleo-cytoplasmic ratio) using the Invitrogen TRIzol reagent, followed by further purification with the Qiagen RNeasy MinElute Cleanup Kit. For each cell line examined three biological replicates were used and each replicate represented either a different passage or different sub-line (which had been cultured independently and without mixing for several passages). The RNA samples were quantified using the NanoDrop ND-1000 UV-Vis spectrophotometer (NanoDrop Technologies) and the quality of the RNA was assessed using Lab-on-a-Chip RNA Pico Chips and the 2100 Bioanalyzer (Agilent Technologies). Samples with electropherograms that showed a size distribution pattern predictive of acceptable microarray assay performance were considered to be of good quality. Two micrograms of total RNA from each sample were amplified and labelled using a single-cycle cDNA synthesis and an *in vitro* transcription cRNA-RNA labelling system (GeneChip one-cycle target labelling and control reagents; Affymetrix). Following successful cRNA amplification, 10 µg of labelled target cRNA was hybridized to rhesus macaque Genome Arrays (Affymetrix) using standard protocols, as described in the GeneChip Expression Analysis manual (http://www.affymetrix.com/support/technical/manual/expression_manual.affx). The rhesus macaque array contains 52,865 probe sets, representing over 20,000 genes. The arrays were scanned using the GeneChip laser scanner (Affymetrix) and image processing, normalization and expression analysis were performed with the Affymetrix GCOS version 1.4 software. MAS-5 statistical analysis was performed to calculate the signal log ratio (SLR) for each probe set comparison, and the gene expression fold changes between two samples were calculated from the SLR using the following formula: fold change = (2^{SLR}). The GCOS 1.4 MAS 5.0 software was used to calculate statistically significant differences in gene expression ($P < 0.002$) between samples.

For the primary microarray comparison analysis (Supplementary Data 3) the log₁₀ of the absolute detected signal for each present 'P' probe set ($P < 0.05$) was plotted in a scatter graph (Supplementary Fig. 8) using Affymetrix GeneChip Operating Software (GCOS) version 1.4. For the correlation value calculations, the microarray data for each individual cell line comparison was filtered to only include probe sets present ($P < 0.05$) in both cell lines. The present probe sets (PP) value details the number of probe sets, post filtering, with a present ($P < 0.05$) signal in both compared cell lines. The correlation value for each cell line comparison was calculated using MAS-5 (Affymetrix microarray suite 5) analysis to calculate the proportion of compared probe sets which demonstrated no significant change 'NC' in gene expression (Supplementary Data 3). Each cell line had three biological replicates and the letter after the cell line name details which replicate was used in the primary microarray comparison analysis.

For the secondary gene-specific analysis, comparison analysis was performed between each of the three control ORMES-10 biological replicates and each of the three somatic donor cell replicates, to give a total of nine somatic-ES cell comparisons (Supplementary Data 4). The following selection criteria were used to identify rhesus somatic-specific genes: (1) genes that were considered to be present ($P < 0.05$) in all three somatic donor cell replicates; and (2) genes that demonstrated statistically significant decrease 'D' in gene expression in the ORMES-10 replicates in all nine comparisons with the somatic donor cell replicates after GCOS comparisons with MAS-5 statistical analysis. A total of 4,998 somatic-specific probe sets were identified in this way (Supplementary Data 5). The following selection criteria were used to identify rhesus ES-cell-specific genes: (1) genes that were considered to be present ($P < 0.05$) in all three ORMES-10 replicates; and (2) genes that demonstrated a statistically significant increase 'I' in gene expression in the ORMES-10 ES cell replicates in all nine comparisons with the somatic donor cell replicates after GCOS comparisons with MAS-5 statistical analysis. A total of 6,178 ES-cell-specific probe sets were identified in this way (Supplementary Data 6). It should be noted that the general approximation when working with large numbers of probe sets is to assume that each probe set represents hybridization to a single gene. However, multiple probe sets can exist for certain genes, so the actual number of genes included in the analysis is significantly lower than the number of probe sets analysed. The somatic-specific and ES-cell-specific genes identified from this comparison analysis were then used to investigate whether the CRES cell lines had successfully downregulated somatic-specific genes (Supplementary Table 4) and successfully upregulated ES-cell-specific genes (Table 2) after comparison analysis with the three somatic donor cell replicates.

For the tertiary stemness gene analysis, stemness genes were derived from our previous transcriptional profiling work with rhesus macaque embryonic stem cell lines¹⁸. These 12 putative stemness genes had the highest average fold change

in gene expression when three undifferentiated biological replicates of ORMES-6 were compared to their *in vitro* differentiated counterparts, and all 12 were significantly upregulated in five different rhesus macaque embryonic stem cell lines examined¹⁸. Comparison analysis was performed between the ORMES and CRES cell line replicates and the donor somatic cell replicates (Supplementary Data 7), and the average fold change increase in gene expression of the 12 stemness genes in the ORMES and CRES cell lines was calculated (Table 3). All of the normalized microarray data sets generated from these studies can be found on the GEO website, as noted above.

Statistical analysis. Non-microarray results were analysed by chi-squared and unpaired *t*-test using Statview Software (SAS Institute, Inc.) with statistical significance set at 0.05. Microarray results were analysed by MAS-5 statistical analysis in Affymetrix GeneChip Operating Software (GCOS) version 1.4.

35. Zelinski-Wooten, M. B., Hutchison, J. S., Hess, D. L., Wolf, D. P. & Stouffer, R. L. Follicle stimulating hormone alone supports follicle growth and oocyte development in gonadotrophin-releasing hormone antagonist-treated monkeys. *Hum. Reprod.* **10**, 1658–1666 (1995).
36. Bavister, B. D. & Yanagimachi, R. The effects of sperm extracts and energy sources on the motility and acrosome reaction of hamster spermatozoa *in vitro*. *Biol. Reprod.* **16**, 228–237 (1977).
37. McKiernan, S. H. & Bavister, B. D. Culture of one-cell hamster embryos with water soluble vitamins: pantothenate stimulates blastocyst production. *Hum. Reprod.* **15**, 157–164 (2000).
38. Kuo, H. C. *et al.* Differentiation of monkey embryonic stem cells into neural lineages. *Biol. Reprod.* **68**, 1727–1735 (2003).
39. Pearson, P. L. *et al.* Report of the committee on comparative mapping. *Cytogenet. Cell Genet.* **25**, 82–95 (1979).
40. Rogers, J. *et al.* An initial genetic linkage map of the rhesus macaque (*Macaca mulatta*) genome using human microsatellite loci. *Genomics* **87**, 30–38 (2006).
41. Domingo-Roura, X., Lopez-Giraldez, T., Shinohara, M. & Takenaka, O. Hypervariable microsatellite loci in the Japanese macaque (*Macaca fuscata*) conserved in related species. *Am. J. Primatol.* **43**, 357–360 (1997).
42. St John, J. C. & Schatten, G. Paternal mitochondrial DNA transmission during nonhuman primate nuclear transfer. *Genetics* **167**, 897–905 (2004).
43. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nature Genet.* **29**, 365–371 (2001).

Innate versus learned odour processing in the mouse olfactory bulb

Ko Kobayakawa^{1*}, Reiko Kobayakawa^{1*}, Hideyuki Matsumoto², Yuichiro Oka¹, Takeshi Imai¹, Masahito Ikawa³, Masaru Okabe³, Toshio Ikeda⁴, Shigeyoshi Itoharu⁴, Takefumi Kikusui⁵, Kensaku Mori² & Hitoshi Sakano¹

The mammalian olfactory system mediates various responses, including aversive behaviours to spoiled foods and fear responses to predator odours. In the olfactory bulb, each glomerulus represents a single species of odorant receptor. Because a single odorant can interact with several different receptor species, the odour information received in the olfactory epithelium is converted to a topographical map of multiple glomeruli activated in distinct areas in the olfactory bulb. To study how the odour map is interpreted in the brain, we generated mutant mice in which olfactory sensory neurons in a specific area of the olfactory epithelium are ablated by targeted expression of the diphtheria toxin gene. Here we show that, in dorsal-zone-depleted mice, the dorsal domain of the olfactory bulb was devoid of glomerular structures, although second-order neurons were present in the vacant areas. The mutant mice lacked innate responses to aversive odorants, even though they were capable of detecting them and could be conditioned for aversion with the remaining glomeruli. These results indicate that, in mice, aversive information is received in the olfactory bulb by separate sets of glomeruli, those dedicated for innate and those for learned responses.

The mouse olfactory system can detect and discriminate diverse odorants using a repertoire of about 1,000 odorant receptor genes¹. Each olfactory sensory neuron (OSN) expresses only one member of the odorant receptor gene family in a monoallelic manner². Furthermore, OSNs expressing the same odorant receptor converge their axons to a specific set of glomeruli in the olfactory bulb³. Thus, odorous information received in the olfactory epithelium is converted to topographical maps of activated glomeruli. On the basis of the expression patterns of zone-specific markers^{4,5}, the olfactory epithelium can be divided into two, non-overlapping areas: a dorsal zone (D zone) and a ventral zone (V zone).

Vertebrate odorant receptor genes are phylogenetically divided into two distinct classes⁶: class I and class II. Class I odorant receptors are expressed exclusively in the D zone of the olfactory epithelium, and OSNs expressing them project their axons to the most antero-dorsal area in the olfactory bulb^{7,8}. In addition to the class I odorant receptors, ~300 class II odorant receptors are also expressed in the D zone, but their corresponding glomeruli reside on the periphery of the class I area in the olfactory bulb. The remaining class II odorant receptors are expressed in the V zone and their glomeruli are found in the ventro-lateral area in the olfactory bulb^{4,9}. Thus, the glomerular map seems to be subdivided into three compartments along the dorso-ventral axis in the olfactory bulb: a dorsal domain for class I odorant receptors (D_I domain), a dorsal domain for class II odorant receptors (D_{II} domain) and a ventral domain for class II odorant receptors (V domain) (Supplementary Fig. 1a). The olfactory bulb can also be divided into distinct domains in other ways, for example, on the basis of the chemical natures and structural features of odorous ligands¹⁰. Because a particular odorant interacts with many different odorant receptor species, multiple sets of glomeruli are activated in different olfactory bulb domains¹¹. However, little is known about how the topographical information in the olfactory bulb is transmitted to and interpreted in

the brain to decode the odour map. To address these questions, we generated two new strains of mutant mice in which the OSNs in a specific area of the olfactory epithelium are ablated.

Generation of the zone-specific depletion mice

We reported previously that the rat *Omacs* gene, encoding olfactory-specific medium-chain acyl-CoA synthetase, is expressed in the olfactory epithelium in a D-zone-specific manner before the onset of odorant receptor gene expression⁵. This was also confirmed in mouse by *in situ* hybridization and immunohistochemistry (Supplementary Fig. 2). We used the D-zone-specific *O-MACS* (also known as *BC048390*) promoter to ablate the D zone OSNs in mice (Δ D) by targeted expression of the diphtheria toxin gene¹². We first generated a knock-in mouse strain in which the coding sequence of the *O-MACS* was replaced with that of the Cre-recombinase gene (*O-MACS*→*cre*). We then crossed this knock-in mouse with another mouse strain carrying the Cre-inducible *lacZ* gene. The *lacZ* reporter, activated by Cre that is driven from the *O-MACS* promoter, was selectively expressed in the D zone of the olfactory epithelium by embryonic day 11. No other tissues, including the brain, expressed *lacZ* (Supplementary Fig. 3). We thus confirmed the D-zone-specific induction of Cre. Next, we crossed the *O-MACS*→*cre* mouse with a knock-in mouse (*Eno2*-STOP-*DTA*) in which the Cre-inducible diphtheria toxin A gene (*DTA*) was introduced into the neuron-specific enolase gene (*Eno2*) locus (Fig. 1a, Δ D). D-zone-specific ablation of OSNs in the olfactory epithelium was confirmed by *in situ* hybridization and by polymerase chain reaction with reverse transcription (RT-PCR) analyses for various odorant receptors (Fig. 1b and Supplementary Fig. 1b, Δ D).

We also generated the class-II-depleted (Δ II) mouse using the olfactory receptor 16 (*MOR23*, also known as *Olfr16*) promoter. We obtained several lines of the *MOR23*→*cre* transgenic mice¹³ and crossed them

¹Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Tokyo 113-0032, Japan. ²Department of Physiology, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan. ³Research Institute for Microbial Diseases, The Osaka University, Osaka 565-0871, Japan. ⁴Laboratory for Behavioral Genetics, Brain Science Institute, RIKEN, Saitama 351-0198, Japan. ⁵Laboratory of Veterinary Ethology, The University of Tokyo, Tokyo 113-8657, Japan.

*These authors contributed equally to this work.

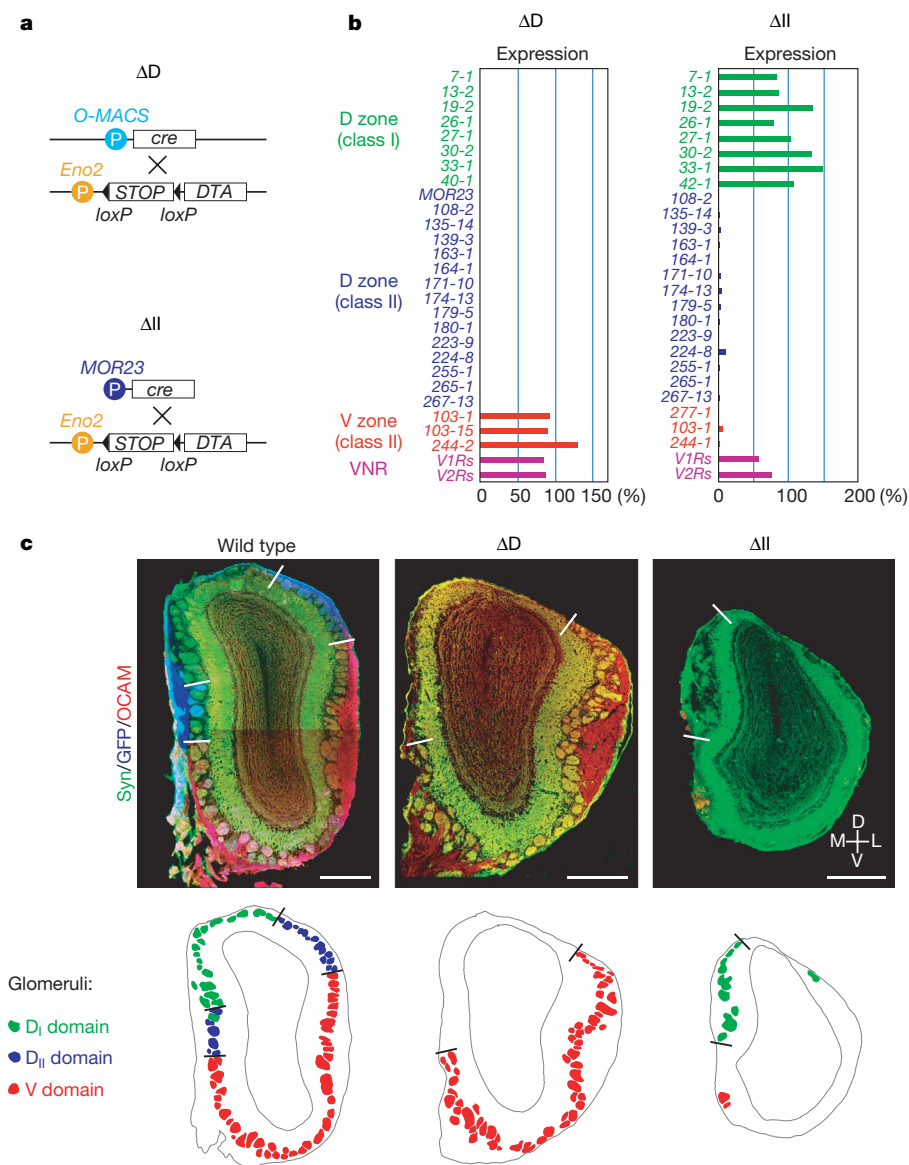


Figure 1 | Odorant receptor gene expression and glomerular map formation in the ΔD and ΔII mutant mice.

a, Plasmid constructs for the D-zone depletion (ΔD) and the class II depletion (ΔII) are schematically illustrated. *DTA* and *cre* are the genes for diphtheria toxin A and Cre recombinase, respectively. **b**, Expression of the odorant receptor and vomeronasal receptor (VNR) genes. Olfactory epithelium sections dissected from the ΔD , ΔII and wild-type mice were analysed by *in situ* hybridization. The number of OSNs expressing a given receptor per unit volume of mutant olfactory epithelium was compared with that of wild-type control (%). In the ΔD mutant, OSNs expressing the D-zone-specific odorant receptor genes, both class I (green) and class II (blue), were ablated. In the ΔII mice, OSNs expressing the class II odorant receptor genes, in the both D (blue) and the V (red) zones, were ablated. Expression of the vomeronasal receptor genes (purple)—*V1Rs* (*V1re8*, *V1rg1* and *V1rj3*) and *V2Rs* (*V2r1b*, *V2r11* and *V2r15*)—was also analysed, and the mean ratios are shown. **c**, Coronal sections of olfactory bulbs. Samples were dissected from the mutant mice (ΔD and ΔII) and the wild-type mouse in which the class II odorant-receptor-expressing OSNs were labelled with GFP. Olfactory bulb sections were stained with anti-synaptotagmin (green, Syn), anti-GFP (blue) and anti-OCAM (also known as NCAM2; red) antibodies. Distributions of glomeruli are schematically shown below the images (D_I, D_{II} and V-domain glomeruli are illustrated in green, blue and red, respectively). Borders for D_I domains (GFP[−]/OCAM[−]) and D_{II} domains (GFP⁺/OCAM[−]) are indicated by solid lines. D, dorsal; L, lateral; M, medial; V, ventral. Scale bars, 500 μ m.

with the *Eno2*-*STOP*-*DTA* mouse to ablate OSNs under the control of the *MOR23* promoter (Fig. 1a, ΔII). With one particular transgenic line, the depletion unexpectedly occurred throughout the entire class II region of the olfactory epithelium (Fig. 1b and Supplementary Fig. 1b, ΔII). *In situ* hybridization and RT-PCR analysis demonstrated that none of the class I odorant receptor genes was affected in this mouse (Fig. 1b and Supplementary Fig. 1b, ΔII), with the exception of two class I genes⁸ expressed in the V zone (data not shown).

Arrangement of glomeruli in the olfactory bulb

We analysed the formation and arrangement of glomeruli in the mutant mice by immunohistochemistry (Fig. 1c and Supplementary Fig. 4). In the ΔD mouse, glomeruli were found only in the postero-ventral region of the olfactory bulb (V domain), leaving the antero-dorsal region (D domain) vacant (Fig. 1c and Supplementary Fig. 4, ΔD). This was also confirmed by imaging of intrinsic signals in the olfactory bulb. In contrast to the wild-type mouse, no odour-evoked activities were detected on the dorsal surface of the olfactory bulb in the ΔD mouse ($n = 3$) (Supplementary Figs 5a, 6 and 7). Despite the complete absence of glomeruli, the D domain of the mutant olfactory bulb showed otherwise normal cytoarchitecture with distinct layers (Supplementary Fig. 8a). Individual mitral cells in the D domain of the mutant olfactory bulb emitted several dendrites into the external plexiform layer, but did not form dendritic terminal tufts, suggesting

that these mitral cells lacked synaptic inputs from OSN axons (Supplementary Fig. 8b, c). V-zone OSN axons did not form ectopic contacts with the D-domain mitral cells in the ΔD olfactory bulb. No glomeruli were found in the D domain in the ΔD mouse during the course of embryonic development (data not shown). These results indicate that mitral cells are specified as D-domain and V-domain subsets before the projection of OSN axons occurs. The D-domain mitral cells seem to be committed to receive olfactory inputs exclusively from D-zone OSNs.

We then analysed the ΔII mouse. In contrast to the ΔD mutant, the ΔII mouse had glomeruli only in the most antero-dorsal part of the D domain (Fig. 1c and Supplementary Fig. 4, ΔII). To determine the boundary of class I and class II areas in the wild-type olfactory bulb, a mouse containing the Cre-inducible green fluorescent protein (*GFP*) gene was crossed with the *MOR23*-*cre* mouse to produce a mouse in which class-II-expressing OSNs are labelled with GFP (class II-GFP). In the class II-GFP mouse, the D_I-domain glomeruli (negative for GFP) and D_{II}-domain glomeruli (positive for GFP) are segregated into two distinct areas in the D domain of the olfactory bulb, even though both types of D-zone OSNs are intermingled within the D zone of the olfactory epithelium⁸ (Fig. 1c and Supplementary Fig. 4, wild type). In the ΔII mouse, glomeruli were found in the D_I domain, and were absent in the class II regions (D_{II} domain and V domain) of the olfactory bulb (Fig. 1c and Supplementary Fig. 4, ΔII).

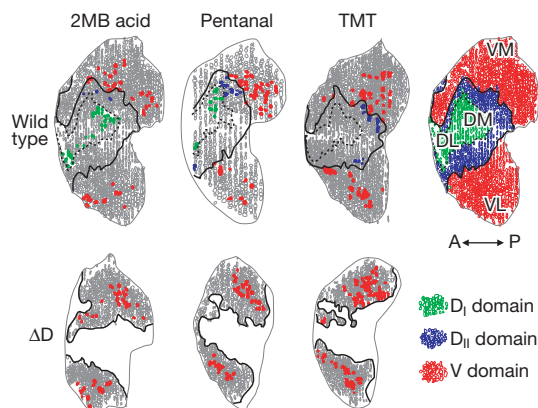


Figure 2 | Odour maps for aversive odorants. Unrolled maps for Zif268 expression. Glomeruli activated by 2MB acid, pentanal and TMT were detected by staining for the immediate-early gene product Zif268. Zif268-positive glomeruli are indicated in the unrolled olfactory bulb maps for wild-type and ΔD mice by three different colours: D_I (green), D_{II} (blue) and V (red). DL, dorso-lateral; DM, dorso-medial; VL, ventro-lateral; VM, ventro-medial.

Odour-activated glomerular maps in the olfactory bulb

A single odorant can interact with several different odorant receptor species and activate multiple sets of glomeruli in distinct areas in the olfactory bulb¹¹. In rats, glomeruli for odorous ligands with similar odour characteristics tend to be clustered in the olfactory bulb¹⁰. This also seems to be true in mice (Supplementary Fig. 6). We analysed odour-activated glomerular maps by optical imaging of intrinsic

signals^{14–16} (Supplementary Figs 5a, 6 and 7) and by expression of Zif268 (also known as Egr1), an immediate-early gene product¹⁷ (Fig. 2 and Supplementary Fig. 9). Three compounds were used: pentanal and 2-methylbutyric (2MB) acid, which are pungent odorants of spoiled foods, and trimethyl-thiazoline (TMT), which is secreted from the anal gland of fox¹⁸ and induces aversive behaviour and fear responses in mice^{19,20}. These odorants activate glomeruli in both D and V domains in the olfactory bulb (Fig. 2 and Supplementary Fig. 9). The 2MB acid activates glomeruli in the D_I and V domains; pentanal activates glomeruli in the D_I , D_{II} and V domains; and TMT activates glomeruli in the D_{II} and V domains. In the ΔD mouse, only V-domain glomeruli were activated (Fig. 2 and Supplementary Figs 5a and 9).

We then analysed the detection thresholds for these odorants by the habituation–dishabituation test²¹ (Supplementary Figs 5b and 10). Comparing the ΔD mouse to the wild-type mouse, the detection thresholds for pentanal and TMT were not affected, whereas it was ten-times higher for 2MB acid. In this regard, it has been reported that, in rats, the most responsive glomerulus for pentanal is located in the V-domain area²². The same seems to be true in mice, not only for pentanal but also for TMT.

The ΔD mice fail to show innate avoidance behaviour

Rodents demonstrate avoidance behaviours towards predators' odorants^{18–20}. They also avoid spoiled smells, for example, aliphatic acids^{19,20}, aliphatic aldehydes²³ and alkyl amines²⁴. In contrast, rodents show attractive behaviours to food smells and conspecific odours²⁵. We performed olfactory preference and avoidance tests with the wild-type and ΔD mutant mice.

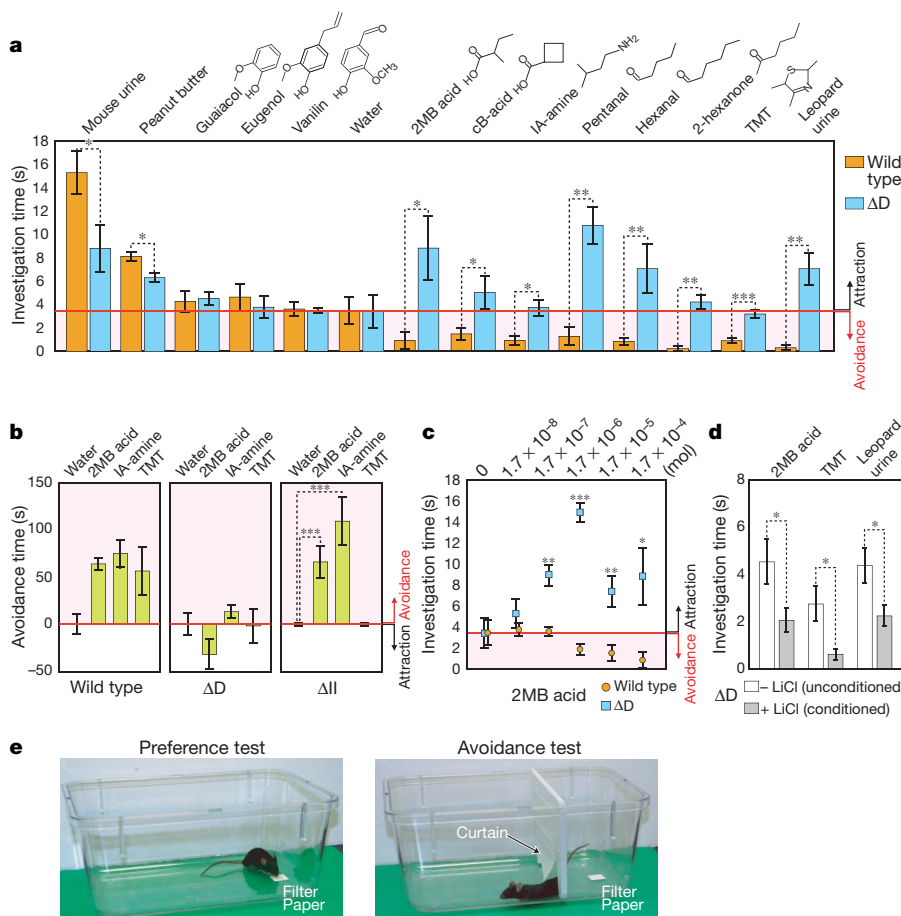


Figure 3 | Recognition tests for innate odour qualities. **a**, Innate olfactory preference tests. The duration for which the mouse investigated the scented filter paper was measured for various odorants. Mean investigation times ($s \pm$ s.e.m.) are shown for each odorant during the 3-min test period with the wild-type (orange) and ΔD (blue) mice. Investigation time for water was used as a criterion for attraction versus avoidance responses (red line). Investigation times less than this criterion (avoidance responses) are marked by the red shaded area. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$. cB acid, cyclobutanecarboxylic acid; IA-amine, iso-amyl amine. **b**, Innate olfactory avoidance tests. Avoidance times were measured in a cage divided into two partitions (1:3) separated by a move-through curtain. The scented filter paper was placed in the smaller partition. During the 3-min test period, the time spent in the larger partition was measured as the avoidance time. Mean avoidance times ($s \pm$ s.e.m.) for test odorants relative to the mean avoidance time for water, which is set to 0, are shown as a bar graph. Triple asterisk, $P < 0.001$. **c**, Preference tests with the serially diluted 2MB acid. Mean investigation times ($s \pm$ s.e.m.) for 2MB acid in various amounts are plotted for the ΔD (blue squares) and wild-type (orange circles) mice. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$ for a significant difference between investigation times of wild-type and ΔD mice for the same amounts of 2MB acid. **d**, Aversive conditioning of the ΔD mice. The ΔD mice were aversively conditioned to the indicated odorants with LiCl injection. Mean investigation times ($s \pm$ s.e.m.) are shown for each odorant during the 3-min test period for aversively conditioned (grey) and unconditioned (white) ΔD mice. Asterisk, $P < 0.05$. **e**, Photos of the preference (left) and avoidance (right) tests.

In the preference test (Fig. 3a and Supplementary Video), filter papers scented with various odorants were presented to the wild-type and ΔD mutant mice. The lengths of time the mouse spent investigating and sniffing the odorants were measured. Investigation times for wild-type mice were approximately 4 s for water (neutral odorant), within 0.3–1.3 s for acids, aldehydes, amines, ketones and predator odours, and 8–15 s for mouse urine and peanut butter. We used the investigation time for water (4 s) as a criterion for avoidance (shorter than 4 s) versus attraction (longer than 4 s) responses. In contrast to the wild-type, the ΔD mice spent a much longer time, 4–10 s, investigating acids, aldehydes, amines, ketones and predator odours. The failure of ΔD mutants to demonstrate avoidance behaviour was also confirmed by the innate avoidance test (Fig. 3b). We performed the innate-olfactory preference test using serially diluted 2MB acid (Fig. 3c). Wild-type mice demonstrated avoidance responses at higher concentrations of 2MB acid, but showed neutral behaviours at lower concentrations. In contrast, ΔD mice showed attractive behaviours at higher concentrations, and neutral responses at lower concentrations.

These results indicate that ΔD mice are able to detect 2MB acid, but fail to show avoidance behaviour. It should be noted that the mouse can be conditioned for aversive reactions by injection of irritating LiCl²⁶. We used 2MB acid, TMT and snow leopard urine in the conditioning test, and found that the investigation time was significantly shortened after conditioning (Fig. 3d). It seems that ΔD mice can be conditioned for aversion, but they lack innately aversive responses.

The D zone is sufficient for innate avoidance behaviour

To examine the discrimination capabilities of the ΔD mouse, we performed an olfactory discrimination test²⁷ using four different pairs of odorants (Fig. 4). In this test, food-restricted animals were trained to associate either of the two related odorants to sugar rewards. After four days of training, the embedded sugar was removed from the test cage, and digging times for the two related odorants were compared within the same animal. It was found that ΔD mutants were able to discriminate 2MB acid, pentanal and (–)carvone from the competing odorants (cyclobutanecarboxylic acid, hexanal and (+)carvone, respectively). Mutant mice could discriminate even subtle differences between enantiomers as well as could wild-type mice. The wild-type mice could not be trained to associate TMT to sugar rewards, because the innate avoidance response was overwhelming. In contrast, the ΔD mouse easily discriminated TMT from eugenol, because the mutant did not avoid TMT.

To examine whether the D domain is sufficient to cause innate avoidance behaviour, we tested another OSN-ablated mutant, the ΔII mouse. The ΔII mouse demonstrated clear avoidance behaviours towards 2MB acid and iso-amyl amine, both of which activate D_I- and V-domain glomeruli (Fig. 3b). Interestingly, the ΔII mice did not avoid TMT, probably because the D-domain glomeruli for TMT are located exclusively in the D_{II} domain, which are absent in the ΔII mice (Fig. 2 and Supplementary Fig. 5a).

The innate fear pathway is not activated in the ΔD mouse

TMT activates the hypothalamic–pituitary–adrenal (HPA) axis in the rat brain, which has a significant role in regulating the adrenocorticotrophic hormone (ACTH) level in response to stress^{18,28}. Neurotoxic lesions of the lateral and basal nuclei of the rat amygdala do not prevent the TMT-induced freezing reaction²⁹. Interestingly, muscimol (an agonist for a neurotransmitter, γ -aminobutyric acid A, GABA_A) does not block the TMT-induced freezing of rats when infused into the amygdala, but does block this reaction when introduced into the bed nucleus of the stria terminalis (BST)³⁰. These observations indicate that the BST is involved in the TMT-induced fear processing. Using the immediate-early gene product Zif268 as a marker, we examined the TMT-induced activation of the BST (Fig. 5a, b). In the wild-type mouse, the BST was strongly activated in the medial aspect (BST-MA), and moderately in the

lateral division (BST-LD). This is consistent with the previous observation that TMT activates the BST-MA, leading to the stimulation of the HPA axis in rats²⁸. In contrast to the wild type, the BST-MA was not activated by TMT in the ΔD mice, although the BST-LD was activated as in wild-type mice. We also examined 2MB acid-induced activation of the BST. In contrast to TMT, 2MB acid activated the BST-LD, but not BST-MA, in both wild-type and ΔD mice (Fig. 5a, b).

The HPA axis in rats, when activated by TMT, increases blood ACTH²⁸. We measured the plasma concentrations of ACTH in both wild-type and ΔD mice (Fig. 5c). ACTH concentrations were found to increase significantly in the wild-type mouse, but not in the ΔD mutant, when TMT was present. These results indicate that the innate fear pathway in the ΔD mouse cannot be activated by TMT. With 2MB acid, the rise of ACTH was very small, if at all, even in the wild-type mouse (Fig. 5c). Although both TMT and 2MB acid cause avoidance responses in the wild-type mouse (Fig. 3a, b), only TMT activates the BST-MA and elevates plasma ACTH (Fig. 5a–c). These results indicate that fear responses induced by TMT and aversive responses to 2MB acid are separately processed in the brain with different neural circuits.

Discussion

In the present study, we generated zone-specific ablation mice, and studied glomerular map formation and odour perception in these mutants. When OSNs in specific olfactory epithelium zones were ablated, the corresponding olfactory bulb areas were devoid of glomerular structures. However, second-order neurons were present in

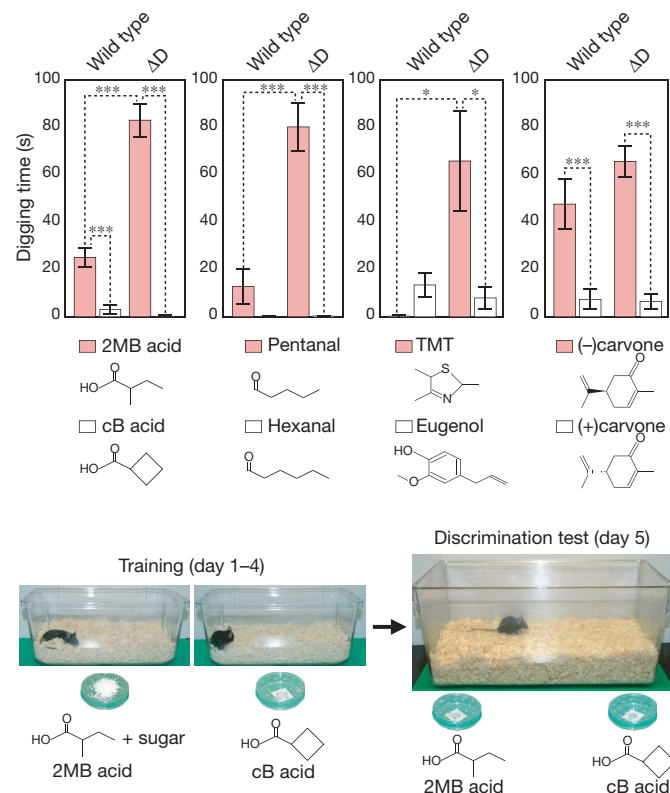


Figure 4 | Discrimination tests with structurally related odorants. The wild-type and ΔD mice were trained for four days to associate the reward (sugar grains) with either of the two related odorants. On day five, the sugar reward was removed from the bed, and digging times were measured for each pair of related odorants. Mean digging times (s) \pm s.e.m. during the 2-min test period are shown as bar graphs for odorants paired with the sugar reward (red bars), and for unpaired odorants (white bars). Asterisk, $P < 0.05$; double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$. Photos of the discrimination test are shown underneath.

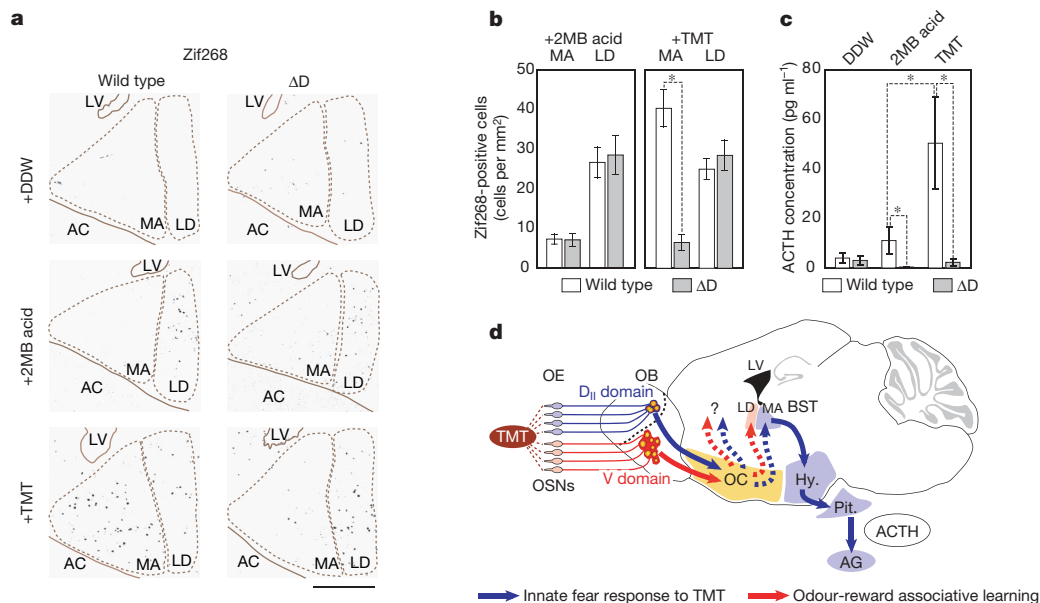


Figure 5 | 2MB-acid- and TMT-evoked neural pathways in the brain.

a, 2MB-acid- and TMT-induced activation of the BST. The BST was activated by 2MB acid and TMT in the wild-type and ΔD mice. Double-distilled water (DDW) was used as a control. Coronal sections of the BST were immunostained with anti-Zif268 antibodies to detect the expression of the immediate-early gene *Zif268*. The lateral ventricle (LV) and the anterior commissure (AC) were used as landmarks. MA, medial division of BST, anterior; LD, lateral division of BST, dorsal. Scale bar, 500 μ m.

b, Quantifications of Zif268-positive cells in the BST. 2MB-acid- and TMT-induced Zif268-positive cells in the BST-MA (MA) and BST-LD (LD) were counted for each brain section shown in **a**. Mean densities of activated cells (Zif268-positive cells per mm²) \pm s.e.m. were compared between the wild-type (white) and ΔD (grey) mice. Asterisk, $P < 0.05$. **c**, Measurements of the 2MB-acid- and TMT-induced rise in plasma ACTH. The plasma ACTH

induced by 2MB acid, TMT or double-distilled water (DDW) was measured in the wild-type and ΔD mice. Average concentrations (pg ml⁻¹) \pm s.e.m. of the plasma ACTH are shown. Asterisk, $P < 0.05$. **d**, A schematic diagram of the neural pathways in the mouse brain. A predator's odorant, TMT, activates two sets of glomeruli, one in the D domain and the other in the V domain of the olfactory bulb. We propose that TMT activates two different neuronal pathways: one for the innate fear response (blue) and the other for the odour-reward association learning (red). For TMT, the D-domain glomeruli activate the olfactory cortex (OC), and, subsequently, the BST-MA, which is thought to activate the HPA axis and increase the plasma ACTH concentration. The V-domain glomeruli are assumed to activate the BST-LD (LD) by means of the olfactory cortex responding to TMT. AG, adrenal gland; Hy., hypothalamus; LV, lateral ventricle; Pit., pituitary gland; OB, olfactory bulb; OE, olfactory epithelium.

the vacant areas in the olfactory bulb. These findings were unexpected, because in other sensory systems, such as retino-tectal projection³¹ and barrel formation³², competing axons eventually occupy vacant projection sites. In the olfactory system, the olfactory bulb may not simply be a projection screen to form a glomerular map, but may, instead, have area-specific functions that are predetermined genetically before the onset of OSN projection.

The present results indicate that D-domain glomeruli and V-domain glomeruli have separate roles in eliciting avoidance behaviours, by sending distinct signals to the brain (Fig. 5d). However, we must be careful in interpreting the data of behavioural tests. It is possible that ΔD mice require more time to recognize the qualities of odorants in the innate preference tests. It has been reported that there is a trade-off between speed and accuracy in rodent olfactory recognition^{33–35}. We should consider the possibility that investigation times for aversive smells could be increased in the ΔD mice, if the V-domain glomeruli are slower to transmit odour information to the brain than the D-domain glomeruli. However, we believe that this is unlikely, because investigation times were decreased for the ΔD mice with attractive odorants (Fig. 3a). Even when the ΔD mice were exposed to TMT or 2MB acid for as long as 10–30 min, they did not demonstrate any increase in the levels of ACTH or Zif268 (Fig. 5a–c). Furthermore, ΔD mice found hidden odours more quickly than wild-type mice (Supplementary Fig. 11). From these results, it is difficult to explain the behavioural phenotypes of the ΔD mice in terms of the speed-accuracy trade-off. It is known that behavioural responses to odorants are affected by their concentrations. There is a possibility that, in the ΔD mice, detection thresholds for odorants are altered. For 2MB acid, the detection threshold was ten times higher for the ΔD mice than for the wild-type mice (Supplementary Figs 5b and 10). However, in the innate preference

test using serially diluted 2MB acid, the ΔD mice were consistently on the attractive side whereas the wild-type mice were on the avoidance side (Fig. 3c). From these results, it is difficult to explain the behavioural abnormalities in ΔD mice only by detection thresholds.

In contrast to the ΔD mutant, ΔII mice showed aversive behaviour towards 2MB acid and iso-amyl amine. It seems that D-domain glomeruli are necessary and sufficient to induce aversive responses towards 2MB acid and iso-amyl amine. Because these odorants activate multiple sets of glomeruli in the D and V domains simultaneously, it was thought that these glomeruli would contribute equally to the processing of odour information in the glomerular map. However, the present study indicates that the mouse main olfactory system may be composed of at least two functional modules: one for innate odour responses and the other for discrimination and associative learning of complex odorous information. Innate aversive/fear responses probably involve genetically programmed neural circuits, like in the mammalian gustatory system^{36,37}, and in the chemosensory systems in *Drosophila melanogaster*^{38,39} and *Caenorhabditis elegans*^{40,41}. Like the immune system, the mouse olfactory system seems to have maintained innate odour responses with hard-wired neural circuits, in parallel with newly acquired discrimination/adaptive circuits.

METHODS SUMMARY

Mice. The ΔD and ΔII mice were generated by crossing the *Eno2-STOP-DTA* mouse with the *O-MACS*→*cre* mouse and *MOR23*→*cre* mouse, respectively. The class II-GFP mouse was generated by crossing the *MOR23*→*cre* mouse with the *ROSA-STOP-GFP* mouse.

Histology. *In situ* hybridization was performed as described^{8,9}. For fluorescence immunohistochemistry, coronal sections of the brains, which were fixed with 4% paraformaldehyde in PBS for 5 min and rinsed in TBS-T (50 mM Tris, 150 mM NaCl and 0.2% Triton-X100), were then blocked with 10% normal goat

serum in TBS-T, and incubated with primary antibodies at 23 °C for 12–16 h. Sections were washed in TBS-T, and incubated with secondary antibodies.

Mapping odour-evoked activities. Adult male mice were habituated to the cage for 2 h, and then filter paper scented with test odors was presented for 30 min. After 1 h, mice were killed and perfused with 4% paraformaldehyde in PBS. Zif268 mapping was performed as described with minor modifications¹⁷.

Innate olfactory preference test. Adult $\Delta\Delta$ and littermate mice ($n = 110$ for each genotype, and $n \geq 6$ for each test) were used in the test. To avoid confounding of data owing to learning, mice were used only once.

After habituating to the experimental environment, mice were transferred to the test cage, and a filter paper scented with a test odorant was introduced. Investigation times for the filter paper during the 3-min test period were measured.

Innate olfactory avoidance test. The $\Delta\Delta$, ΔII and wild-type littermates (adult male) ($n = 26$ for each genotype, and $n \geq 6$ for each test) were used in the test. The test cage was divided into two compartments (1:3) with a partition curtain. A filter paper scented with test odorant was introduced into the smaller compartment. Times (seconds) spent in the larger compartment were measured.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 August; accepted 14 September 2007.

Published online 7 November 2007.

- Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
- Serizawa, S., Miyamichi, K. & Sakano, H. One neuron–one receptor rule in the mouse olfactory system. *Trends Genet.* **20**, 648–653 (2004).
- Mombaerts, P. et al. Visualizing an olfactory sensory map. *Cell* **87**, 675–686 (1996).
- Yoshihara, Y. et al. OCAM: a new member of the neural cell adhesion molecule family related to zone-to-zone projection of olfactory and vomeronasal axons. *J. Neurosci.* **17**, 5830–5842 (1997).
- Oka, Y. et al. O-MACS, a novel member of the medium-chain acyl-CoA synthetase family, specifically expressed in the olfactory epithelium in a zone-specific manner. *Eur. J. Biochem.* **270**, 1995–2004 (2003).
- Zhang, X. & Firestein, S. The olfactory receptor gene superfamily of the mouse. *Nature Neurosci.* **5**, 124–133 (2002).
- Zhang, X. et al. High-throughput microarray detection of olfactory receptor gene expression in the mouse. *Proc. Natl Acad. Sci. USA* **101**, 14168–14173 (2004).
- Tsuboi, A., Miyazaki, T., Imai, T. & Sakano, H. Olfactory sensory neurons expressing class I odorant receptors converge their axons on an antero-dorsal domain of the olfactory bulb in the mouse. *Eur. J. Neurosci.* **23**, 1436–1444 (2006).
- Miyamichi, K., Serizawa, S., Kimura, H. M. & Sakano, H. Continuous and overlapping expression domains of odorant receptor genes in the olfactory epithelium determine the dorsal/ventral positioning of glomeruli in the olfactory bulb. *J. Neurosci.* **25**, 3586–3592 (2005).
- Mori, K., Takahashi, Y. K., Igarashi, K. M. & Yamaguchi, M. Maps of odorant molecular features in the Mammalian olfactory bulb. *Physiol. Rev.* **86**, 409–433 (2006).
- Malnic, B., Hirono, J., Sato, T. & Buck, L. B. Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999).
- Gogos, J. A., Osborne, J., Nemes, A., Mendelsohn, M. & Axel, R. Genetic ablation and restoration of the olfactory topographic map. *Cell* **103**, 609–620 (2000).
- Imai, T., Suzuki, M. & Sakano, H. Odorant receptor-derived cAMP signals direct axonal targeting. *Science* **314**, 657–661 (2006).
- Rubin, B. D. & Katz, L. C. Optical imaging of odorant representations in the mammalian olfactory bulb. *Neuron* **23**, 499–511 (1999).
- Uchida, N., Takahashi, Y. K., Tanifuji, M. & Mori, K. Odor maps in the mammalian olfactory bulb: domain organization and odorant structural features. *Nature Neurosci.* **3**, 1035–1043 (2000).
- Gurden, H., Uchida, N. & Mainen, Z. F. Sensory-evoked intrinsic optical signals in the olfactory bulb are coupled to glutamate release and uptake. *Neuron* **52**, 335–345 (2006).
- Inaki, K., Takahashi, Y. K., Nagayama, S. & Mori, K. Molecular-feature domains with posterodorsal–anteroventral polarity in the symmetrical sensory maps of the mouse olfactory bulb: mapping of odourant-induced Zif268 expression. *Eur. J. Neurosci.* **15**, 1563–1574 (2002).
- Vernet-Maury, E., Polak, E. H. & Demael, A. Structure–activity relationship of stress-inducing odorants in the rat. *J. Chem. Ecol.* **10**, 1007–1018 (1984).
- Hebb, A. L. et al. Odor-induced variation in anxiety-like behavior in mice is associated with discrete and differential effects on mesocorticolimbic cholecystokinin mRNA expression. *Neuropsychopharmacology* **27**, 744–755 (2002).
- Hebb, A. L. et al. Brief exposure to predator odor and resultant anxiety enhances mesocorticolimbic activity and enkephalin expression in CD-1 mice. *Eur. J. Neurosci.* **20**, 2415–2429 (2004).
- Ferguson, J. N. et al. Social amnesia in mice lacking the oxytocin gene. *Nature Genet.* **25**, 284–288 (2000).
- Johnson, B. A. & Leon, M. Modular glomerular representations of odorants in the rat olfactory bulb and the effects of stimulus concentration. *J. Comp. Neurol.* **422**, 496–509 (2000).
- Wood, R. W. & Coleman, J. B. Behavioral evaluation of the irritant properties of formaldehyde. *Toxicol. Appl. Pharmacol.* **130**, 67–72 (1995).
- Dielensberg, R. A. & McGregor, I. S. Defensive behavior in rats towards predatory odors: a review. *Neurosci. Behav. Rev.* **25**, 597–609 (2001).
- Nyby, J., Kay, E., Bean, N. J., Dahinden, Z. & Kerchner, M. Male mouse attraction to airborne urinary odors of conspecific and to food odors; effects of food deprivation. *J. Comp. Psychol.* **99**, 479–490 (1985).
- Kay, E. & Nyby, J. LiCl aversive conditioning has transitory effects on pheromonal responsiveness in male house mice (*Mus domesticus*). *Physiol. Behav.* **52**, 105–113 (1992).
- Schellinck, H. M., Forestell, C. A. & LoLordo, V. M. A simple and reliable test of olfactory learning and memory in mice. *Chem. Senses* **26**, 663–672 (2001).
- Day, H. E., Masini, C. V. & Campeau, S. The pattern of brain *c-fos* mRNA induced by a component of fox odor; 2,5-dihydro-2,4,5-trimethylthiazoline (TMT), in rats, suggests both systemic and processive stress characteristics. *Brain Res.* **1025**, 139–151 (2004).
- Wallece, K. J. & Rosen, J. B. Neurotoxic lesions of the lateral nucleus of the amygdala decrease conditioned fear but not unconditioned fear of a predator odor: comparison with electrolytic lesions. *J. Neurosci.* **21**, 3619–3627 (2001).
- Fendt, M., Endres, T. & Apfelbach, R. Temporary inactivation of the bed nucleus of the stria terminalis but not the amygdala blocks freezing induced by trimethylthiazoline, a component of fox feces. *J. Neurosci.* **23**, 23–28 (2003).
- Horder, T. J. Retention, by fish optic nerve fibres regenerating to new terminal sites in the tectum, of ‘chemospecific’ affinity for their original sites. *J. Physiol. (Lond.)* **6**, 53P–55P (1971).
- Van der Loos, H. & Woolsey, T. A. Somatosensory cortex: structural alterations following early injury to sense organs. *Science* **179**, 395–398 (1973).
- Uchida, N. & Mainen, Z. F. Speed and accuracy of olfactory discrimination in the rat. *Nature Neurosci.* **6**, 1224–1229 (2003).
- Abraham, N. M. et al. Maintaining accuracy at the expense of speed: stimulus similarity defines odor discrimination time in mice. *Neuron* **44**, 865–876 (2004).
- Rinberg, D., Koulakov, A. & Gelperin, A. Speed–accuracy tradeoff in olfaction. *Neuron* **51**, 351–358 (2006).
- Zhao, G. Q. et al. The receptors for mammalian sweet and umami taste. *Cell* **115**, 255–266 (2003).
- Huang, A. L. et al. The cells and logic for mammalian sour taste detection. *Nature* **442**, 934–938 (2006).
- Suh, G. S. et al. A single population of olfactory sensory neurons mediates an innate avoidance behaviour in *Drosophila*. *Nature* **431**, 854–859 (2004).
- Stockinger, P., Kvitsiani, D., Rotkopf, S., Tirian, L. & Dickson, B. J. Neural circuitry that governs *Drosophila* male courtship behavior. *Cell* **121**, 795–807 (2005).
- Troemel, E. R., Kimmel, B. E. & Bargmann, C. I. Reprogramming chemotaxis responses: sensory neurons define olfactory preferences in *C. elegans*. *Cell* **91**, 161–169 (1997).
- Tobin, D. et al. Combinatorial expression of TRPV channel proteins defines their sensory functions and subcellular localization in *C. elegans* neurons. *Neuron* **35**, 307–318 (2002).
- Shirshak, D. R. et al. Codon-improved Cre recombinase (iCre) expression in the mouse. *Genesis* **32**, 19–26 (2002).
- Feng, G. et al. Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron* **28**, 41–51 (2000).
- Serizawa, S. et al. Mutually exclusive expression of odorant receptor transgenes. *Nature Neurosci.* **3**, 687–693 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Sprengel for the pBS-iCre vector, J. R. Sanes for the *Thy1-YFP-G* mouse and Tama Zoo for the snow leopard’s urine. We also thank Y. Maruyama and A. Kawai for technical assistance. We are grateful to P. Mombaerts, J. Ngai, T. Yamamori and members of our laboratory for comments. This work was supported by the CREST program of the Japan Science and Technology Agency; the Mitsubishi Foundation; and the Special Promotion Research Grant from the Ministry of Education, Science and Culture of Japan. R.K. is supported by the PRESTO program of the Japan Science and Technology Agency.

Author Contributions K.K. and R.K. planned and performed most of the experiments. Optical imaging of intrinsic signals and mitral cell labelling were done by H.M. and K.M. Y.O. participated in the initial characterization of the *O-MACS*→*cre* knock-in mouse. T. Imai generated the *MOR23*→*cre* transgenic mouse. M.I. and M.O. helped generate the *O-MACS*→*cre* mouse. T. Ikeda and S.I. generated the *Eno2-STOP-DTA* mouse. T.K. advised on the behavioral analysis. H.S. supervised the project. The manuscript was written by K.K., R.K. and H.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.S. (sakano@mail.ecc.u-tokyo.ac.jp).

METHODS

Mice. To generate the *O-MACS*→*cre* mice, a 1.3-kilobase (kb) fragment just upstream of the start codon of the *O-MACS* gene and a 6.9-kb fragment containing exon 2 and exon 3 of *O-MACS* were amplified by PCR from the mouse genome, and subcloned into pBluescript (Stratagene), containing the herpes simplex virus thymidine kinase (*hsv-tk*) gene. The SV40-polyA was amplified by PCR from pEGFP-N1 (Clontech), and integrated downstream of the codon-improved Cre recombinase (*iCre*) gene in pBlue.iCre⁴². The resulting iCre-SV40 polyA sequence was integrated in-frame into the ATG site of the *O-MACS* gene. The phosphoglycerate kinase 1 (pgk)-Neo-polyA sequence flanked by two FLP recognition targets (FRTs) derived from pPGK-neo-FRT (Gene Bridges), was inserted downstream of the iCre-polyA. The resulting targeting construct was transfected into embryonic stem cells, D3. Generation of the *MOR23*→*cre* transgenic mice has been described¹³. The *Eno2-STOP-DTA* mouse (T. Ikeda and S.I., unpublished observations) will be described elsewhere. The *ROSA-STOP-lacZ* (stock number 3309) and *ROSA-STOP-GFP* (stock number 4077) reporter mice were purchased from The Jackson Laboratory. The *Thyl1-YFP-G*⁴³ mouse (Supplementary Fig. 8) was provided by J. R. Sanes. The *O-MACS*→*cre* knock-in mouse is in a mixed background of C57BL6 and 129Svj. Both *MOR23*→*cre* and *Eno2-STOP-DTA* mice are of the C57BL6 background. For behavioural analyses, the wild-type littermates were used as controls.

Histology. *In situ* hybridization was performed as described^{8,9}. Antisense RNA probes for the odorant receptor genes^{8,9} and *O-MACS*⁵ were prepared as described. Other probes were prepared from the following open reading frame sequences by PCR: *V2r1b* (nucleotides 1,696–2,551), *V2r11* (also known as *Vmn2r89*, 1,599–2,941), *V2r15* (42–911), *V1re9* (710–1,396), *V1rj3* (279–821), *V1rg1* (100–732) and *Reln* (960–1,920). For fluorescence immunohistochemistry, coronal sections of the brains from animals, which were perfused with 4% paraformaldehyde, were fixed with 4% paraformaldehyde in PBS for 5 min and rinsed in TBS-T (50 mM Tris, 150 mM NaCl and 0.2% Triton-X100) three times. Samples were then blocked with 10% normal goat serum in TBS-T, and incubated with primary antibodies at room temperature (23 °C) overnight (12–16 h). Sections were washed in TBS-T three times, and incubated with secondary antibodies conjugated to Cy3 (1:200, Jackson ImmunoResearch), Alexa Fluor-488 (1:200, Molecular Probes) or Alexa Fluor-647 (1:200, Molecular Probes). The following primary antibodies were used: O-CAM (1:1,000)⁴, GFP (1:100, Nacalai Tesque), synaptotagmin (1:500, Chemicon), Zif268 (1:500, Santa Cruz Biotechnology), GABA (1:2,000, Sigma) and tyrosine hydroxylase (1:200, Chemicon). For labelling of mitral cells, the dorsal surface of the olfactory bulb was surgically exposed, and a glass micropipette (filled with 2% neurobiotin in 1 M NaCl) was inserted vertically into the olfactory bulb. Positive current was applied for labelling the neuron (300–400 nA at 50% duty cycle of 7-s pulses for 30 min). The coronal olfactory bulb sections were dissected from the labelled animal and stained with the ABC kit (Vector Laboratories). Anatomical analyses of the olfactory epitheliums and olfactory bulbs of ΔD and ΔII mice will be described elsewhere (R.K. *et al.*, manuscript in preparation). X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactoside) staining was performed as described⁴⁴.

Polymerase chain reaction with reverse transcription. DNase-treated total RNA extracted from the olfactory epithelium with the RNeasy kit (Qiagen) was used to prepare random-primed complementary DNA with Superscript III (Invitrogen). PCR primers have been described^{8,9}.

Mapping odour-evoked activities. Adult male mice were habituated to the cage for 2 h, and then filter paper scented with one of the following odorants was presented for 30 min: 2MB acid (870 μ M), *n*-pentanal (370 μ M) or TMT (1,200 μ M). After 1 h, mice were killed and perfused with 4% paraformaldehyde in PBS. For Zif268 mapping, the published method¹⁷ was modified. Consecutive coronal sections of the class II-GFP (wild-type) and ΔD mice ($n = 1$ for each condition) after the exposure to odorants, were immunostained with anti-Zif268 or anti-OCAM antibodies. The sections were also immunostained with anti-GFP and anti-synaptotagmin antibodies to identify the boundary of the D_I and D_{II} domains in the wild-type olfactory bulb. Zif268-positive and -negative glomeruli were mapped on the unrolled map. Boundaries between the D_{II} and V domains were identified by OCAM. For the Zif268 mapping in the BST, coronal sections were immunostained with anti-Zif268 antibodies. Zif268-positive cells were counted for both ΔD and wild-type mice ($n = 3$ for each genotype).

Unrolled maps of the olfactory bulb. Unrolled maps were constructed as described^{15,17}. In brief, a smooth line was traced along the centre of the glomerular layer of immunostained coronal sections. The line was flattened by opening at the most ventro-lateral point. For glomerular arrangement (Fig. 2 and Supplementary Figs 4, 5a, 6 and 7), D_I -, D_{II} - and V-domain glomeruli were mapped on the flattened lines. For mapping the odour-evoked activities (Fig. 2), Zif268-positive and -negative glomeruli were mapped along the flattened lines. The unrolled map was generated by aligning the flattened traces of

consecutive sections using the dorso-medial edge of the glomerular layer as a reference. For Supplementary Figs 6 and 7, the imaged region was determined with reference to the dye-injected points, and superimposed onto the map.

Optical imaging of intrinsic signals. Imaging and marker-dye injections were performed on adult, male, class II-GFP ($n = 2$) and ΔD ($n = 3$) mice as described¹⁵. Consecutive coronal sections of olfactory bulb from imaged animals were immunostained with anti-OCAM, anti-GFP and anti-synaptotagmin antibodies. Dorsal views of glomeruli in the D_I domain (GFP⁺/OCAM⁺), D_{II} domain (GFP⁺/OCAM⁺) and V domain (OCAM⁺) were constructed and superimposed onto the optical imaging signals using dye-injected points as references.

Olfactory habituation–dishabituation test. The test was performed as described²¹ with minor modifications using adult littermates ($n = 59$ for ΔD and $n = 61$ for wild type; $n \geq 6$ for each test). Mice were habituated to the cage (20 \times 15 \times 13 cm), and then a filter paper (2 \times 2 cm) with 20 μ l of distilled water was presented for 3 min. This was repeated three times with 1-min intervals. On the fourth trial, a filter paper with the indicated amount of the test odorant was presented for 3 min. Investigation times during the 3-min test period were measured on the third and fourth trials. The mouse behaviour was recorded with a digital video camera (30 frames per second, 640 \times 480 pixels) for the analysis. We defined ‘an investigation’ as a nasal contact with the filter paper within a 1 mm distance.

Innate olfactory preference test. Adult ΔD and littermate mice ($n = 110$ for each genotype, and $n \geq 6$ for each test) were used in the test. To avoid confounding of data owing to learning, mice were used only once. To habituate to the experimental environment, mice were placed individually in a cage that was identical to the test cage (20 \times 15 \times 13 cm). After 30 min of habituation, mice were transferred to a new cage. This habituation was repeated four times for each animal. Soon after the habituation, mice were transferred to the test cage, and a filter paper (2 \times 2 cm) scented with a test odorant was introduced. Investigation times for the filter paper during the 3-min test period were measured. The mouse behaviour was recorded with a digital video camera for the analysis. Odorants used were: guaiacol (158 μ M), eugenol (128 μ M), vanillin (64 μ M), 2MB acid (174 μ M), cyclobutanecarboxylic acid (206 μ M), iso-amyl amine (16.8 μ M), *n*-pentanal (74 μ M), hexanal (156 μ M), 2-hexanone (158 μ M, Tokyo chemical industry), TMT (314 μ M, Pherotech), peanut butter (10% w/v, 40 μ l), urine of female mice (20 μ l) and snow leopard urine (20 μ l). Aversion conditioning with LiCl was adapted from the published procedure⁴² as follows. The ΔD mouse ($n = 36$, and $n = 6$ for each test) was first exposed to test odorants for 3 min as described for the innate olfactory preference test. For aversive conditioning, 0.5 M LiCl solution was immediately injected (20 ml kg⁻¹) into the peritoneal cavity. After 2 days, the mice were subjected to the olfactory preference test with the same odorants, and investigation times (seconds) were measured.

Innate olfactory avoidance test. The ΔD , ΔII and wild-type littermates (adult male) ($n = 26$ for each genotype, and $n \geq 6$ for each test) were used in the test. To avoid confounding owing to learning, each mouse was used only once in the test. The test cage (20 \times 15 \times 13 cm) was divided into two compartments (1:3) with a partition curtain. Mice could move freely between the two compartments. The mice were habituated to the cage for ~10 min, and then a filter paper scented with 20 μ l test odorant was introduced into the smaller compartment. Times (seconds) spent in the larger compartment were measured during the 3-min test period. Avoidance times were recorded as the time spent in the larger compartment for each odorant subtracted by the time spent for distilled water. The mouse behaviour was recorded with a digital video camera for the analysis. Odorants used were 2MB acid (8.7 M), iso-amyl amine (8.4 M) and TMT (15.7 M).

Olfactory discrimination test. The test was performed as described²⁷ with minor modifications. The ΔD and wild-type littermates (adult male) ($n = 24$ for each genotype, $n = 6$ for each test) were tested. The mice were food-restricted to maintain 80–85% of their free feeding weights and trained to associate one of the two related odorants with the sugar reward for 4 days. During the training, the mice received four 10-min trials a day: two trials for an odour paired with sugar reward and two for the unpaired odour. On day five, each test odorant was placed independently, without sugar, in the cage (26 \times 40 \times 18 cm) with bedding (~4,000 cm³). The mouse behaviour was recorded with a digital video camera for the analysis. The time (seconds) spent digging for each odorant was measured during the 2-min test period. Odorant concentrations were: 2MB acid (8.7 M), cyclobutanecarboxylic acid (10.3 M), *n*-pentanal (3.7 M), hexanal (7.8 M), eugenol (6.4 M), (–)carvone (6.3 M, Tokyo chemical industry), (+)carvone (6.2 M, Tokyo chemical industry) and TMT (15.7 M). For each odorant, 20 μ l was used.

ACTH assay. The ACTH assay was adapted from the published procedure²⁸. The ΔD and wild-type littermates (10–12-week-old male) were used ($n = 24$ for each genotype, $n = 8$ for each analysis). The mice were housed individually to avoid

the elevation of plasma ACTH owing to fighting. Mice were maintained with 12:12 h light:dark schedule (lights turned on at 9:00). Odour presentation and blood sampling were performed between 10:00 and 11:00. Filter paper, scented with either double-distilled water or TMT (77.6 μ l each), was quietly dropped into the cage with great care not to excite the animal. After exposure to the odorant for 10 min, the mice were decapitated quickly (within 15 s of removing them from their cages), and blood samples were collected. Plasma ACTH concentrations were measured using the ACTH ELISA kit (MD Biosciences).

Statistics. Data are presented as mean \pm s.e.m. Statistical analysis was performed using a Student's *t*-test. The criterion for statistical significance is $P < 0.05$ in all cases.

Human CtIP promotes DNA end resection

Alessandro A. Sartori¹, Claudia Lukas², Julia Coates¹, Martin Mistrik², Shuang Fu³, Jiri Bartek², Richard Baer³, Jiri Lukas² & Stephen P. Jackson¹

In the S and G2 phases of the cell cycle, DNA double-strand breaks (DSBs) are processed into single-stranded DNA, triggering ATR-dependent checkpoint signalling and DSB repair by homologous recombination. Previous work has implicated the MRE11 complex in such DSB-processing events. Here, we show that the human CtIP (RBBP8) protein confers resistance to DSB-inducing agents and is recruited to DSBs exclusively in the S and G2 cell-cycle phases. Moreover, we reveal that CtIP is required for DSB resection, and thereby for recruitment of replication protein A (RPA) and the protein kinase ATR to DSBs, and for the ensuing ATR activation. Furthermore, we establish that CtIP physically and functionally interacts with the MRE11 complex, and that both CtIP and MRE11 are required for efficient homologous recombination. Finally, we reveal that CtIP has sequence homology with Sae2, which is involved in MRE11-dependent DSB processing in yeast. These findings establish evolutionarily conserved roles for CtIP-like proteins in controlling DSB resection, checkpoint signalling and homologous recombination.

DSBs are highly cytotoxic lesions induced by ionizing radiation and certain anti-cancer drugs. They also arise when replication forks encounter other lesions and are generated as intermediates during meiotic recombination¹. Cells possess two main DSB repair mechanisms: non-homologous end-joining and homologous recombination. Whereas non-homologous end-joining predominates in G0/G1 and is error-prone, homologous recombination is restricted to S and G2, when sister chromatids allow faithful repair^{2–4}. Homologous recombination is initiated by resection of DSBs to generate single-stranded DNA (ssDNA) that binds RPA. A ssDNA–RAD51 nucleoprotein filament then forms to initiate strand invasion. ssDNA–RPA also recruits ATR, triggering ATR-dependent checkpoint signalling by the protein kinase CHK1 (also known as CEHK1) (ref. 5). Because DSB resection is largely restricted to S and G2, both homologous recombination and checkpoint signalling by ATR are subject to cell-cycle control^{6–8}.

A factor implicated in DSB resection is the MRE11–RAD50–NBS1/ NBN (MRN) complex, which binds DNA ends, possesses exo- and endo-nuclease activities and functions in triggering DNA-damage checkpoints^{9,10}. Here, we show that human CtIP physically and functionally interacts with MRN. CtIP—initially identified as an interactor for CtBP¹¹ and the tumour suppressor proteins RB1 (ref. 12) and BRCA1 (refs 13, 14)—is recruited to DNA damage and complexes with BRCA1 to control the G2/M DNA-damage checkpoint^{15–17}. We reveal that CtIP also promotes ATR activation and homologous recombination by cooperating with MRN to mediate DSB resection. On the basis of these findings and our identification of sequence homologies between CtIP and Sae2—a protein implicated in mitotic and meiotic DSB processing in *Saccharomyces cerevisiae*^{18–22}—we conclude that these proteins are functional homologues.

CtIP affects cellular responses to DSBs

To investigate CtIP function, we examined how its depletion affected clonogenic survival of human U2OS cells, following their treatment with DNA-damaging agents. To circumvent possible effects arising from CtIP's involvement in controlling cell-cycle progression, we employed acute (1 h) drug treatments rather than continuous drug

exposure. Short interfering (si)RNA-mediated depletion of CtIP caused hypersensitivity towards the topoisomerase I inhibitor camptothecin and the topoisomerase II inhibitor etoposide (Fig. 1a, b) but did not alter cell-cycle distribution profiles and only modestly decreased the proportion of replicating cells (Supplementary Fig. 1a). Furthermore—and consistent with the major cytotoxic lesions caused by camptothecin and etoposide being DSBs arising during S phase²³—the effects of CtIP depletion were largely abrogated by treating cells with the DNA-replication inhibitor aphidicolin (Fig. 1a, b). These data therefore suggested that CtIP depletion impairs cell survival when DSBs are generated during S phase. Also, because DSBs generated in S phase by camptothecin are repaired by homologous recombination²⁴, these results suggested that CtIP promotes homologous recombination. Congruent with CtIP being unnecessary for non-homologous end-joining, its depletion caused only weak hypersensitivity towards bleocin, which generates DSBs at all cell-cycle stages (Fig. 1c, and Supplementary Fig. 1b).

We next assessed the impact of CtIP on phosphorylations mediated by the protein kinases ATM and ATR. In response to DSBs, ATM phosphorylates the checkpoint kinase CHK2 (also known as CHEK2) and histone H2AX (also known as H2AFX; generating a protein species termed γ H2AX)^{25,26}. In contrast, ATR activation by ssDNA elicits CHK1 phosphorylation⁵. As shown in Fig. 1d, camptothecin triggered γ H2AX formation and the generation of a slower-migrating, hyper-phosphorylated species of CtIP (CtIP hyper-phosphorylation was demonstrated by phosphatase treatments and a phospho-specific antibody; Supplementary Fig. 1c, d). Furthermore, both CtIP and H2AX phosphorylation were prevented when aphidicolin was co-administered with camptothecin. Etoposide also triggered modification of H2AX and CtIP; however, in this case, CtIP phosphorylation but not H2AX phosphorylation was abrogated by aphidicolin (Fig. 1e). These data are in accord with etoposide being able to generate DSBs throughout the cell cycle, and suggest that CtIP phosphorylation occurs most effectively in S phase.

CtIP depletion did not affect H2AX phosphorylation triggered by either camptothecin or etoposide, revealing that CtIP is not required for DNA-lesion generation by these drugs (Fig. 1d, e). Similarly,

¹The Wellcome Trust and Cancer Research UK Gurdon Institute, and Department of Zoology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK. ²Institute of Cancer Biology and Centre for Genotoxic Stress Research, Danish Cancer Society, Strandboulevarden 49, DK-2100 Copenhagen, Denmark. ³Institute for Cancer Genetics, Department of Pathology, Columbia University, New York, New York 10032, USA.

camptothecin-induced phosphorylations of CHK2 and SMC1 were unaffected by CtIP depletion, indicating that ATM activation still occurs when CtIP is absent (Fig. 1f, and data not shown). In contrast, CtIP depletion diminished camptothecin-induced CHK1 phosphorylation on Ser 345 and Ser 317 (Fig. 1f, and data not shown). Notably, CtIP depletion markedly impaired CHK1 phosphorylation at late time-points but not at early time-points after continuous camptothecin exposure (Supplementary Fig. 1e). This suggested that CtIP is not required for initial CHK1 phosphorylation triggered by replication stress induced by trapped DNA topoisomerase I molecules, but is needed for signalling of camptothecin-induced replication-dependent DSBs, which, as for ionizing-radiation-induced DSBs^{7,27}, might require resection to elicit ATR activation. We reasoned that if this

was the case, CtIP might influence hyper-phosphorylation of the amino-terminal region of RPA2 by ATM, ATR and DNA-PK (also known as PRKDC) in the context of replication-associated DSBs^{28,29}. Indeed, camptothecin-induced RPA2 phosphorylation—as measured by a shift in RPA2 electrophoretic mobility and by phospho-specific antibodies against modified RPA2 serines 4 and 8—was markedly impaired by CtIP depletion, whereas DSB formation, as measured by γ H2AX formation, was unaffected (Fig. 1g). Similarly, CtIP depletion impaired RPA hyper-phosphorylation but not CHK2 or SMC1 phosphorylation in response to etoposide (data not shown). Together, these data indicated that CtIP is required neither for DSB generation by topoisomerase inhibitors nor for DSB detection and signalling by ATM. Instead, CtIP is specifically needed for efficient activation and/or propagation of ATR-mediated signalling.

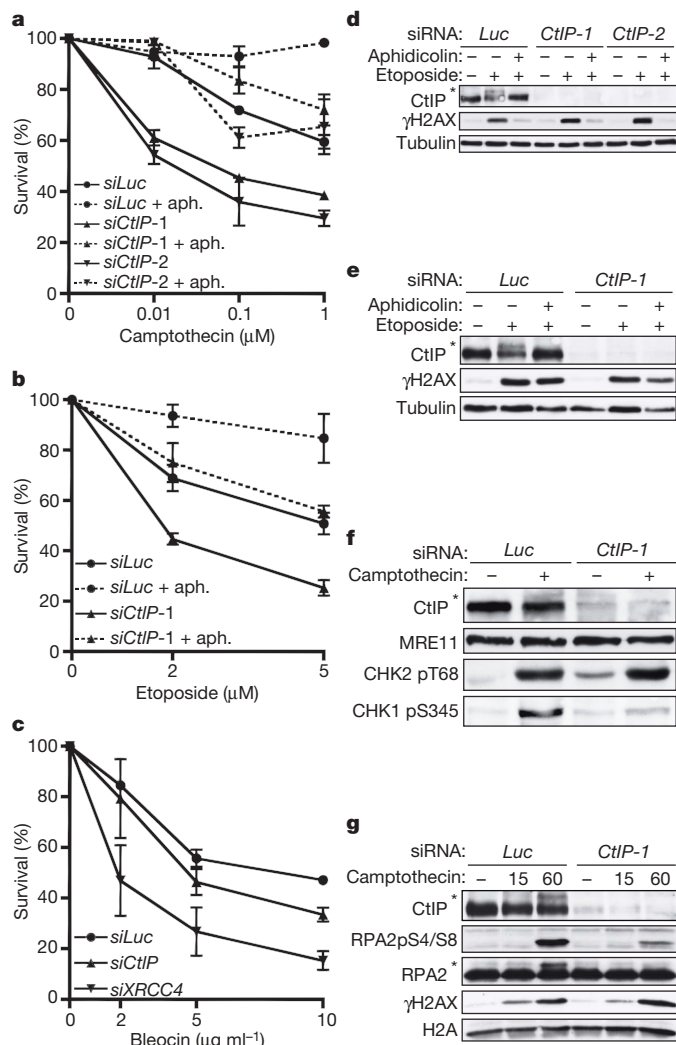


Figure 1 | CtIP depletion causes hypersensitivity to DSB-inducing agents. **a–c**, CtIP downregulation causes replication-dependent camptothecin and etoposide hypersensitivity, and weak bleocin hypersensitivity. U2OS cells were pre-incubated with aphidicolin (aph.) where indicated, then treated for 1 h with camptothecin, etoposide or bleocin. *CtIP-1* and *CtIP-2* are two independent siRNAs. siRNAs against Luciferase (*siLuc*) and XRCC4 (*siXRCC4*) were used as internal controls and are described in more detail in Supplementary Methods. Survival data represent mean \pm s.e.m. from ≥ 3 independent experiments. **d, e**, Aphidicolin suppresses camptothecin- and etoposide-induced CtIP phosphorylation. Extracts from cells downregulated for luciferase or CtIP and treated with camptothecin or etoposide in the presence or absence of aphidicolin were immunoblotted as indicated. **f, g**, CtIP depletion impairs CHK1 and RPA phosphorylation but not CHK2 phosphorylation after camptothecin treatment. H2A, histone H2A; RPA2pS4/S8 phosphorylation sites are defined in the text. Asterisks in **d–g**: hyper-phosphorylated CtIP and RPA2.

CtIP promotes ATR recruitment in S/G2

When we generated DSB-containing tracks in human U2OS cell nuclei by laser micro-irradiation³⁰, CtIP was recruited to the damage with kinetics slower than that of γ H2AX formation (Fig. 2a).

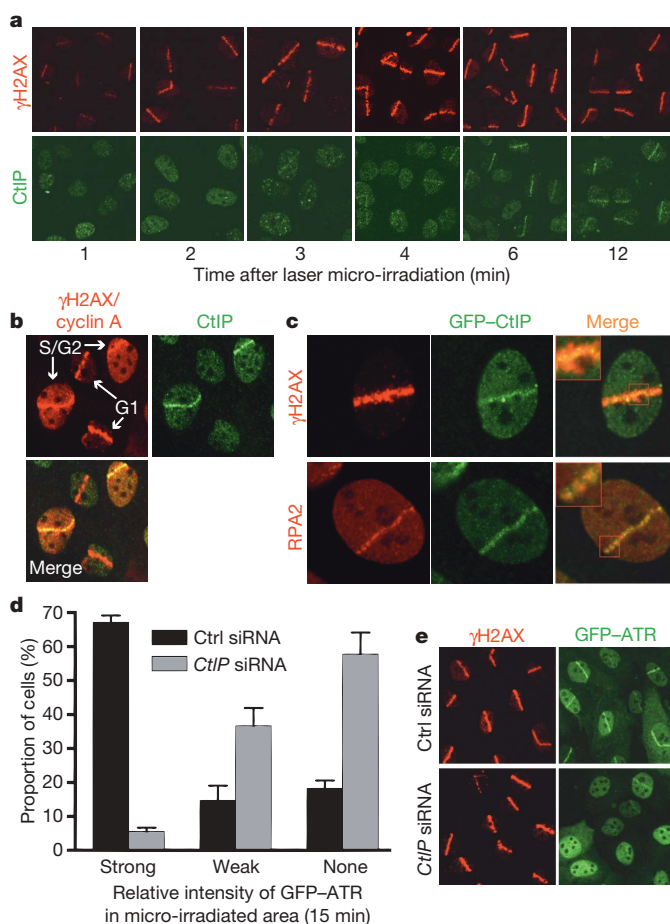


Figure 2 | CtIP associates with sites of DNA damage in S/G2 phase and promotes ATR recruitment to DSBs. **a, b**, CtIP recruitment to laser-induced DSBs occurs in S/G2 phase. Cells were stained for endogenous γ H2AX, CtIP or cyclin A. All cells have local γ H2AX signals but only S/G2-phase cells have pan-nuclear cyclin A staining. **c**, GFP-CtIP co-localizes with RPA-ssDNA. Insets: higher magnifications. **d**, CtIP downregulation impairs ATR recruitment to DNA damage. GFP-ATR-expressing cells were treated, micro-irradiated and monitored (Supplementary Fig. 2c). GFP-ATR DSB tracks were manually scored 15 min after micro-irradiation for control (Ctrl) and CtIP-depleted cells (296 and 264 cells, respectively). Data represent mean \pm s.e.m. from two experiments. **e**, CtIP depletion does not affect γ H2AX formation. Cells from **d** were fixed 15 min after micro-irradiation and immunostained. Wherever used in experiments, images to evaluate cellular responses to laser damage were acquired with a 40 \times /1.2 numerical aperture C-Apochromat objective (Zeiss); the nuclear diameter corresponds to approximately 10 μ m.

Notably, whereas γ H2AX tracks formed in all irradiated cells, only a subset showed strong CtIP recruitment, indicating that CtIP recruitment might be affected by cell-cycle status. Indeed, significant CtIP redistribution to laser tracks was only observed in cells staining positive for cyclin A, indicating that CtIP responds predominantly to DSBs in S and G2 (Fig. 2b).

In the above studies, we noted that—similar to what has been reported for RPA and ATR³¹—CtIP stripes were narrower than those of γ H2AX. Indeed, when we irradiated a U2OS cell line stably expressing green fluorescent protein (GFP)-tagged CtIP (GFP-CtIP; Supplementary Fig. 2a), GFP-CtIP did not cover the entire γ H2AX-modified chromatin region but localized to smaller compartments that resembled those occupied by RPA (Fig. 2c). Consistent with this, endogenous CtIP co-localized with GFP-ATR (Supplementary Fig. 2b). Because ATR localizes to RPA foci, and because we had found that CtIP promotes ATR activation, we used a cell line expressing functional GFP-ATR⁷ to examine whether CtIP is needed for ATR recruitment. Indeed, whereas control cells exhibited strong GFP-ATR recruitment to laser tracks, this response was markedly attenuated in CtIP-depleted cells (Fig. 2d, and Supplementary Fig. 2c; note that the siRNA *CtIP-1* is used here and thereafter). Importantly, parallel immunostaining studies on fixed cells revealed that CtIP depletion impaired ATR recruitment but not γ H2AX formation, implying that CtIP does not affect the extent of laser-induced DNA damage (Fig. 2e). Nevertheless, some residual ATR recruitment was observed in CtIP-depleted cells, indicating either that CtIP stimulates but is not absolutely required for ATR recruitment, or that it is essential for ATR recruitment but that CtIP siRNA-depletion was not complete.

CtIP facilitates the generation of ssDNA

Recruitment of the ATR-ATRIP complex to DNA damage is mediated by interactions with RPA bound to ssDNA^{5,32}. To see whether CtIP might affect ssDNA generation, we analysed DSB-induced RPA focus formation. Strikingly, CtIP depletion dramatically impaired RPA recruitment to laser-induced DNA damage (Fig. 3a; as shown in Supplementary Fig. 2d, this defect was ameliorated when cells stably expressed siRNA-resistant GFP-CtIP, indicating that the GFP-CtIP fusion is functional and that the phenotypes caused by the siRNA are CtIP-dependent). Similarly, camptothecin exposure generated large numbers of RPA foci in control cells (~40% showed this pattern, reflecting the proportion of cells in S phase), whereas CtIP-depleted cells generally lacked detectable RPA foci despite having normal γ H2AX induction (Fig. 3b). Moreover, by using an anti-BrdU antibody staining technique that

only detects DNA in single-stranded form, we found that camptothecin triggered substantial ssDNA formation in control cells but not in CtIP-depleted cells (Fig. 3c, and Supplementary Fig. 3a, b). These data therefore established that CtIP promotes ssDNA formation.

CtIP interacts with MRN and promotes DNA repair

MRN promotes the processing of DSBs to generate RPA-coated ssDNA, which is needed for ATR recruitment and ensuing CHK1 phosphorylation^{7,33,34}. Having found that CtIP was also required for these events, and because BRCA1 and CtIP co-purify with RAD50 in gel-filtration analyses¹⁵, we examined whether MRN and CtIP interact. Thus, we carried out immunoprecipitations from human HeLa cell nuclear extracts with antisera raised against the C- or N-terminal regions of CtIP. As shown in Fig. 4a, MRN and BRCA1 were detected in CtIP immunoprecipitates. Although most CtIP was recovered by such immunoprecipitations, BRCA1 was only inefficiently retrieved, possibly reflecting the G2-specific nature of the CtIP-BRCA1 interaction¹⁶. Furthermore, only a proportion of MRN was recovered in CtIP immunoprecipitates, suggesting that there might be different populations of MRN, with only a subset CtIP-associated. The interaction between CtIP and MRN was quite stable, surviving washing in 0.5 M NaCl, and persisted in the presence of ethidium bromide, indicating that it was not DNA-mediated (data not shown). Conversely, CtIP was retrieved in RAD50 and MRE11 immunoprecipitates (Fig. 4b, and data not shown). The CtIP-MRN interaction was unaffected by camptothecin treatment and took place when BRCA1 was depleted (Fig. 4b). Furthermore, when we used extracts from human HCC1937 cells, CtIP was found in RAD50 immunoprecipitates and vice versa (Supplementary Fig. 4a). Because HCC1937 cells express low levels of functionally impaired BRCA1 bearing a mutation in its C-terminal tandem BRCT domain that precludes its interaction with CtIP¹⁶, these data implied that BRCA1 does not bridge the CtIP-MRN interaction. Indeed, CtIP co-immunoprecipitated with RAD50 from mixtures containing only purified, recombinant CtIP and MRN, demonstrating a direct interaction between the two factors (Supplementary Fig. 4b).

To characterize the CtIP-MRN interaction further, we expressed in bacteria various regions of CtIP fused to glutathione S-transferase (GST) and assessed their ability to retrieve MRN complexes from HeLa nuclear extracts (Fig. 4c). Whereas MRN did not bind to N-terminal regions of CtIP containing a putative coiled-coil or the BRCA1 interaction motif¹⁶ (lanes 4 and 5), it bound efficiently when GST was fused to full-length CtIP or to a region comprising the C-terminal 108 residues of CtIP (lanes 3 and 7; see Supplementary Fig. 6a for a schematic view of the CtIP fragments). Furthermore,

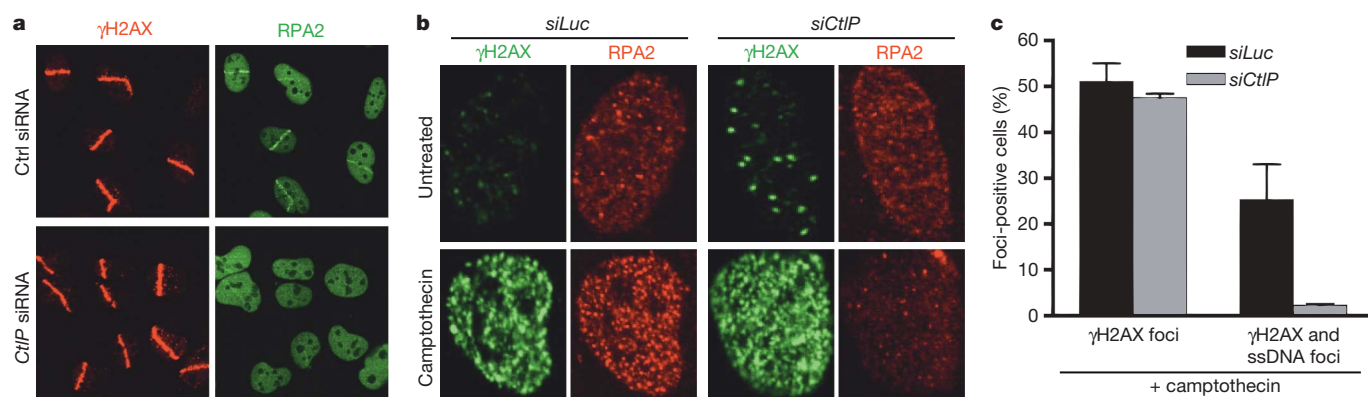


Figure 3 | CtIP depletion impairs DSB resection. **a**, **b**, CtIP is required for RPA recruitment to laser- and CPT-induced DSBs. Cells treated with control or CtIP siRNA were either micro-irradiated and 30 min later co-immunostained for γ H2AX and RPA2, or were treated with camptothecin for 1 h and immunostained. **c**, CtIP depletion impairs ssDNA formation. After siRNA treatment, cells were tested for camptothecin-induced ssDNA

formation by a non-denaturing BrdU staining procedure (see Methods and text for details). For each camptothecin-treated sample, >100 cells were counted and the percentage exhibiting γ H2AX foci, or both γ H2AX and ssDNA foci, was determined (Supplementary Fig. 3a, b). Data represent the mean \pm s.e.m. from two independent experiments.

when we used purified GST-fusions of CtIP (expressed in bacteria or insect cells), they bound directly to purified, baculovirus-expressed human MRE11–RAD50 (Supplementary Fig. 4c,d). However, whereas GST-fusions containing the CtIP C-terminal region interacted with MRE11–RAD50 (Supplementary Fig. 4c), a GST–CtIP fusion lacking the CtIP C-terminal region still interacted with MRE11–RAD50 (Supplementary Fig. 4d). These data therefore indicated that CtIP and MRN interact directly and that there is more than one point-of-contact between the two factors.

Because both CtIP and MRE11 are needed for effective DSB resection, we speculated that CtIP might affect nuclease activities associated with the MRE11 complex^{10,35–37}. Thus, we expressed CtIP in insect cells, purified it to homogeneity (Supplementary Fig. 4b) and used it together with purified MRE11–RAD50 in an endonuclease assay with closed-circular single-stranded PhiX174 DNA (Fig. 4d). As shown previously³⁸, MRE11–RAD50 exhibited nuclease activity in the presence of manganese (lane 12) but not in the presence of magnesium (lane 5). In contrast, CtIP did not show detectable nuclease activity with either metal cofactor (lanes 3 and 10). Strikingly, when CtIP was combined with MRE11–RAD50, this produced endonuclease activity in the presence of magnesium (lane 6) and produced more endonuclease activity in the presence of manganese than was exhibited by MRE11–RAD50 alone (compare lanes 12 and 13). Heat denaturation of CtIP abolished its stimulatory effects and ATP was required neither for MRE11–RAD50 activity nor for the stimulatory effect of CtIP (data not shown). Notably, when we carried out MRE11–RAD50-dependent exonuclease assays with short (50 bp) radiolabelled oligonucleotide substrates, however, we did not detect CtIP-dependent effects (data not shown). We therefore speculate that CtIP mainly promotes endo- but not exonucleolytic activities in conjunction with the MRE11 complex.

Given that both CtIP and MRN regulate DSB resection, we speculated that they might facilitate homologous recombination. To address this, we used a U2OS cell line bearing one or two copies of an integrated homologous recombination reporter composed of two differentially mutated GFP genes oriented as direct repeats. Transient expression of I-SceI endonuclease in such cells generates a DSB that, when repaired by gene conversion, results in a functional GFP gene, the expression of which can be assessed by flow cytometry (Supplementary Fig. 5a)³⁹. Significantly, depletion of CtIP or MRE11 decreased homologous recombination frequencies to levels similar to those achieved by depleting the key homologous recombination protein RAD51 (Fig. 4e, and Supplementary Fig. 5b). Furthermore, depleting CtIP and MRE11 together did not decrease homologous recombination efficiency further than was achieved by CtIP or MRE11 depletion alone. These findings therefore revealed that both CtIP and MRE11 promote homologous recombination, and furthermore suggested that they do so through a common mechanism.

CtIP has homology to *Saccharomyces cerevisiae* Sae2

Prompted by the above findings, we used regions of CtIP in PSI-BLAST-based database searches⁴⁰. Although most regions of CtIP identified related proteins only in higher eukaryotes, we identified proteins in a diverse range of eukaryotes that show homology with the C-terminal 108 amino acid region of CtIP (Supplementary Fig. 6a). Most strikingly, the only *S. cerevisiae* protein identified in this search was Sae2, a protein that genetically interacts with yeast Mre11 to govern camptothecin sensitivity and DSB processing (Fig. 5a, and Supplementary Fig. 6a)^{18–22,41,42}. To address the potential functional importance of the CtIP/Sae2 homology region, we generated a U2OS cell line stably expressing an siRNA-resistant GFP-tagged CtIP derivative (residues 1–789) lacking the C-terminal region. Analysis of

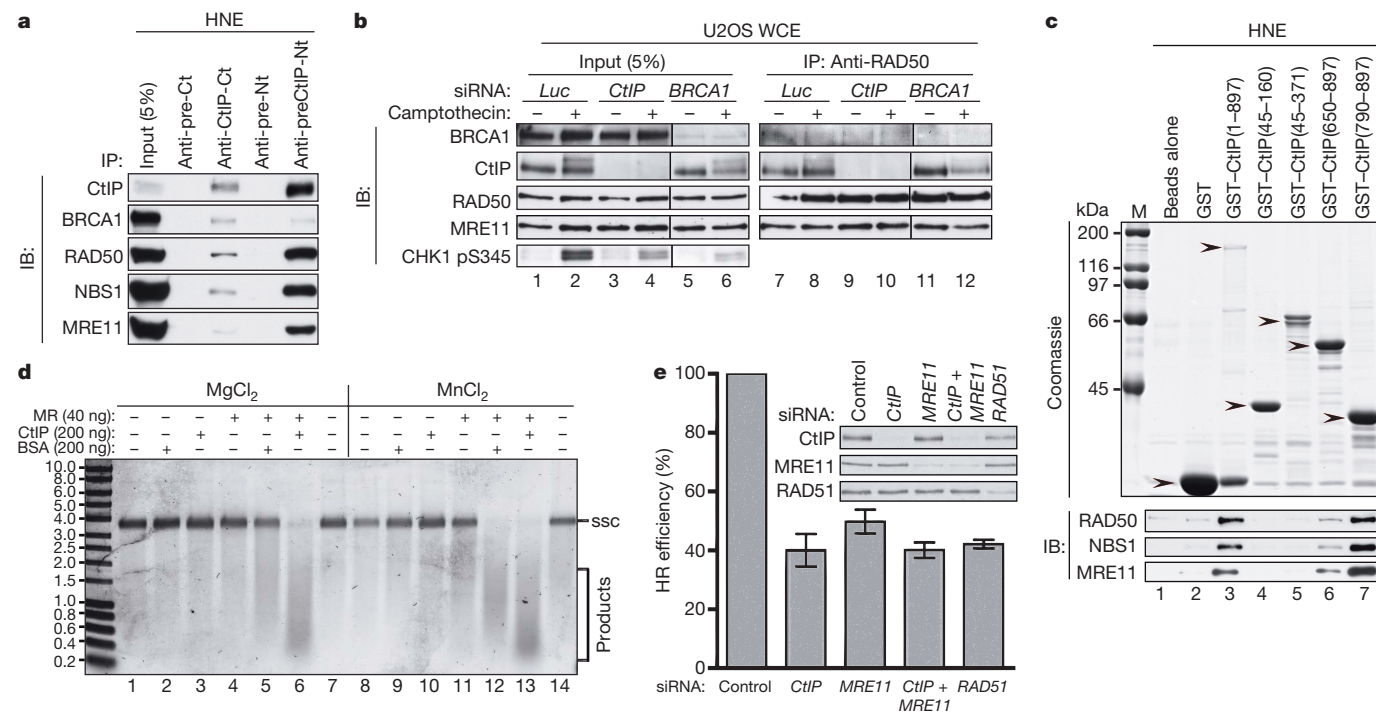


Figure 4 | CtIP interacts with MRN and promotes homologous recombination. **a**, MRN co-immunoprecipitates with CtIP. HeLa nuclear extract (HNE) was immunoprecipitated (IP) with pre-immune (pre) or anti-CtIP antibodies (see Supplementary Methods and text for details) and analysed by immunoblotting (IB). **b**, CtIP–MRN interaction after DNA damage and after BRCA1-depletion. Where indicated, cells were treated with 1 μ M CPT for 1 h. Whole-cell extracts (WCE) were immunoblotted directly or after immunoprecipitation. **c**, The CtIP C terminus binds MRN. Bacterially expressed fusions (arrowheads) were tested for binding MRN in

HNE by immunoblotting. **d**, CtIP stimulates MRE11–RAD50-dependent nuclease activity. PhiX174 substrate was incubated with MRE11–RAD50 (MR; 40 ng), BSA (200 ng) or CtIP (200 ng) in 5 mM MgCl₂ or MnCl₂, run on an agarose gel and stained with SYBR Gold. A series of size markers (in kb) is indicated on the left. ssC, circular ssDNA. **e**, CtIP or MRE11 downregulation impairs homologous recombination (see Methods and text for details). Data represent the mean \pm s.e.m. from four independent experiments. HR, homologous recombination.

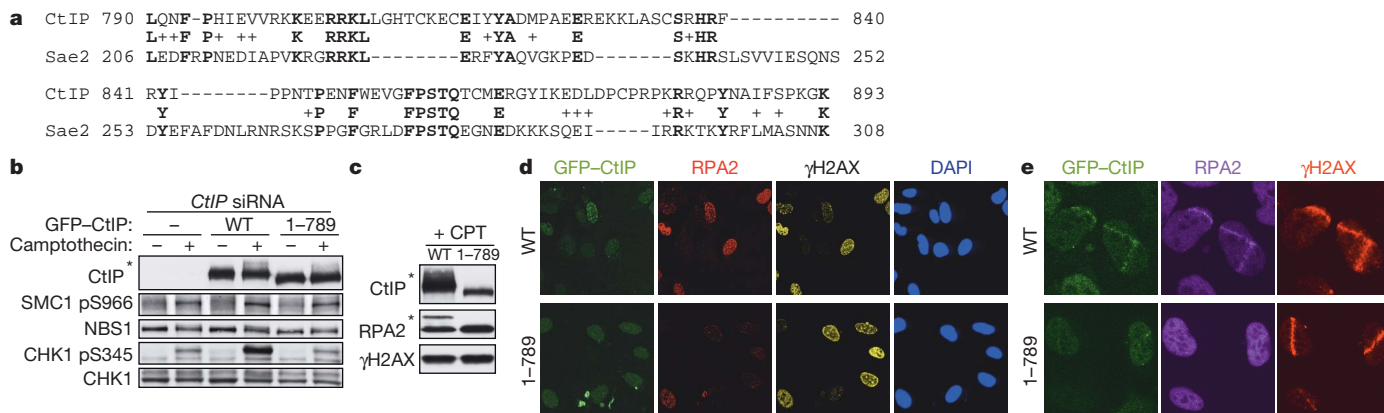


Figure 5 | Function and evolutionary conservation of the CtIP C terminus. **a**, Alignment of CtIP and Sae2 arising from BLAST-searches with the CtIP C terminus (790–897). **b–e**, Deletion of the CtIP C terminus impairs CtIP function. Three days after siRNA transfection, U2OS cells stably expressing GFP-tagged siRNA-resistant wild-type CtIP (WT) or a deletion mutant

(1–789) were either treated with 1 μM CPT for 1 h and analysed by immunoblotting or co-immunostaining, or were micro-irradiated and 30 min later co-immunostained as indicated. Asterisks in **b** and **c**: hyperphosphorylated CtIP and RPA2.

these cells under conditions where the endogenous CtIP protein was downregulated established that, unlike the wild-type protein, the CtIP truncation mutant did not promote efficient CHK1 phosphorylation, RPA2 hyper-phosphorylation or RPA focus formation on camptothecin treatment (Fig. 5b–d). Similarly, wild-type CtIP promoted the formation of RPA-coated ssDNA in laser micro-irradiation studies, whereas the CtIP C-terminal truncation mutant did not (Fig. 5e; a similar proportion of wild-type and mutant cells stained positive for cyclin A, ruling out cell-cycle effects—data not shown). As anticipated from previous experiments (Supplementary Fig. 4d), both wild-type CtIP and the truncation mutant could be co-immunoprecipitated with RAD50 (Supplementary Fig. 6b). Taken together, these data therefore indicated that the conserved C-terminal region of CtIP is functionally important.

Discussion

We have shown that CtIP facilitates DSB resection and ssDNA formation, thus activating the ATR/CHK1 axis of the DNA-damage response and DNA repair by homologous recombination. Consequently, cells depleted of CtIP show hypersensitivity towards DSB-inducing agents, particularly those that generate DSBs specifically in S phase. Furthermore, we have established that CtIP functionally interacts with the MRN complex. Taken together with the fact that MRN also promotes DSB resection, ATR signalling and homologous recombination, these findings identify CtIP as a critical regulator of the checkpoint signalling and DNA repair functions of the MRN complex.

We have also found that the CtIP C terminus is highly conserved in CtIP orthologues in other organisms and is required for CtIP function in human cells. Strikingly, this domain shows homology with the C terminus of *S. cerevisiae* Sae2, which genetically cooperates with the yeast Mre11 complex to promote DSB resection. We therefore conclude that CtIP and Sae2 are probably functional counterparts and that proteins related to these factors will turn out to control DSB resection, ssDNA-mediated checkpoint signalling and homologous recombination in diverse eukaryotes. Efficient DSB processing is restricted to S and G2 phases, and requires CDK and ATM kinase activity⁷. It is therefore noteworthy that CtIP and Sae2 are phosphorylated by ATM and the yeast counterpart of ATR^{43–45}, respectively, and that the CtIP/Sae2 homology region contains conserved potential CDK and ATM/ATR target sites. This raises the exciting possibility that phosphorylation of these sites regulates DSB resection. In addition, CtIP interacts with and is ubiquitinated by BRCA1 (refs 13, 14, 16, 17), and has been identified as an RB1 interaction-partner¹², suggesting that these interactions might control the fate of DSBs during the cell cycle. Finally, it is significant that CtIP governs

responses to commonly used anti-cancer agents and that CtIP alterations have been reported in some cancer cells, thus highlighting CtIP and its interaction with MRN as potential targets for drug discovery^{13,46,47}.

METHODS SUMMARY

HeLa, HCC1937, U2OS and U2OS-derived cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and standard antibiotics. Data for survival curves were generated by colony formation assays. In brief, U2OS cells were transfected with siRNA (see Supplementary Methods) and treated with DNA-damage-inducing drugs. After 1 h, the drug was removed and cells were left for 10–14 days at 37 °C to allow colonies to form. Colonies were stained with 0.5% crystal violet/20% ethanol and counted. Where indicated, cells were pre-incubated with aphidicolin (10 μM) for 90 min, then treated with the specified drug and aphidicolin for 1 h. A U2OS-derived cell line stably expressing GFP-ATR was described previously⁷. The siRNA-resistant silent wild-type GFP-CtIP construct was generated by sub-cloning the CtIP complementary DNA into the pEGFP-C1 expression plasmid (BD Biosciences Clontech) and changing three nucleotides in the CtIP-1 siRNA targeting region by using a QuikChange site-directed mutagenesis kit (Stratagene), as previously described¹⁶. The plasmid expressing the CtIP mutant lacking the C terminus (1–789) was generated by changing residues 790 and 791 in the wild-type GFP-CtIP construct to two stop-codons. For generation of cell lines stably expressing siRNA-resistant GFP-tagged wild-type and mutant CtIP, U2OS cells were transfected with the appropriate constructs and, following antibiotic selection, resistant clones were tested for expression and nuclear localization of the transgene-product by immunofluorescence microscopy. To detect ssDNA by microscopy, cells were cultivated for 24 h in medium supplemented with 10 μM BrdU before camptothecin treatment and, after fixation, immunostained with an anti-BrdU antibody (see Methods) without any preceding DNA denaturation or nuclease treatment⁴⁸. Laser micro-irradiation was performed as described previously^{30,31}. Recombinant Flag-GST-CtIP-6H was isolated from baculovirus-infected Sf9 cells as described previously⁴⁹. Recombinant MRE11–RAD50 and MRE11–RAD50–NBS1 (MRN) complex were gifts from T. Paull.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 April; accepted 5 October 2007.

Published online 28 October 2007.

- Wyman, C. & Kanaar, R. DNA double-strand break repair: all's well that ends well. *Annu. Rev. Genet.* **40**, 363–383 (2006).
- Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. Mechanism and regulation of human non-homologous DNA end-joining. *Nature Rev. Mol. Cell Biol.* **4**, 712–720 (2003).
- West, S. C. Molecular views of recombination proteins and their control. *Nature Rev. Mol. Cell Biol.* **4**, 435–445 (2003).
- Sung, P. & Klein, H. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nature Rev. Mol. Cell Biol.* **7**, 739–750 (2006).

5. Zou, L. & Elledge, S. J. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* **300**, 1542–1548 (2003).
6. Ira, G. *et al.* DNA end resection, homologous recombination and DNA damage checkpoint activation require CDK1. *Nature* **431**, 1011–1017 (2004).
7. Jazayeri, A. *et al.* ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks. *Nature Cell Biol.* **8**, 37–45 (2006).
8. Aylon, Y., Liefshitz, B. & Kupiec, M. The CDK regulates repair of double-strand breaks by homologous recombination during the cell cycle. *EMBO J.* **23**, 4868–4875 (2004).
9. Lavin, M. F. The Mre11 complex and ATM: a two-way functional interaction in recognising and signaling DNA double strand breaks. *DNA Repair (Amst.)* **3**, 1515–1520 (2004).
10. D'Amours, D. & Jackson, S. P. The Mre11 complex: at the crossroads of DNA repair and checkpoint signalling. *Nature Rev. Mol. Cell Biol.* **3**, 317–327 (2002).
11. Schaeper, U., Subramanian, T., Lim, L., Boyd, J. M. & Chinnadurai, G. Interaction between a cellular protein that binds to the C-terminal region of adenovirus E1A (CtBP) and a novel cellular protein is disrupted by E1A through a conserved PLDL motif. *J. Biol. Chem.* **273**, 8549–8552 (1998).
12. Fusco, C., Reymond, A. & Zervos, A. S. Molecular cloning and characterization of a novel retinoblastoma-binding protein. *Genomics* **51**, 351–358 (1998).
13. Wong, A. K. C. *et al.* Characterization of a carboxy-terminal BRCA1 interacting protein. *Oncogene* **17**, 2279–2285 (1998).
14. Yu, X., Wu, L. C., Bowcock, A. M., Aronheim, A. & Baer, R. The C-terminal (BRCT) domains of BRCA1 interact *in vivo* with CtIP, a protein implicated in the CtBP pathway of transcriptional repression. *J. Biol. Chem.* **273**, 25388–25392 (1998).
15. Greenberg, R. A. *et al.* Multifactorial contributions to an acute DNA damage response by BRCA1/BARD1-containing complexes. *Genes Dev.* **20**, 34–46 (2006).
16. Yu, X. & Chen, J. DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of BRCA1 C-terminal domains. *Mol. Cell Biol.* **24**, 9478–9486 (2004).
17. Yu, X., Fu, S., Lai, M., Baer, R. & Chen, J. BRCA1 ubiquitinates its phosphorylation-dependent binding partner CtIP. *Genes Dev.* **20**, 1721–1726 (2006).
18. McKee, A. H. & Kleckner, N. A general method for identifying recessive diploid-specific mutations in *Saccharomyces cerevisiae*, its application to the isolation of mutants blocked at intermediate stages of meiotic prophase and characterization of a new gene *SAE2*. *Genetics* **146**, 797–816 (1997).
19. Prinz, S., Amon, A. & Klein, F. Isolation of *COM1*, a new gene required to complete meiotic double-strand break-induced recombination in *Saccharomyces cerevisiae*. *Genetics* **146**, 781–795 (1997).
20. Rattray, A. J., McGill, C. B., Shafer, B. K. & Strathern, J. N. Fidelity of mitotic double-strand-break repair in *Saccharomyces cerevisiae*: a role for *SAE2/COM1*. *Genetics* **158**, 109–122 (2001).
21. Lobachev, K. S., Gordenin, D. A. & Resnick, M. A. The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell* **108**, 183–193 (2002).
22. Clerici, M., Mantiero, D., Lucchini, G. & Longhese, M. P. The *S. cerevisiae* Sae2 protein promotes resection and bridging of double-strand break ends. *J. Biol. Chem.* **280**, 38631–38638 (2005).
23. D'Arpa, P., Beardmore, C. & Liu, L. F. Involvement of nucleic acid synthesis in cell killing mechanisms of topoisomerase poisons. *Cancer Res.* **50**, 6919–6924 (1990).
24. Saleh-Gohari, N. *et al.* Spontaneous homologous recombination is induced by collapsed replication forks that are caused by endogenous DNA single-strand breaks. *Mol. Cell Biol.* **25**, 7158–7169 (2005).
25. Shiloh, Y. ATM and related protein kinases: safeguarding genome integrity. *Nature Rev. Cancer* **3**, 155–168 (2003).
26. Bakkenist, C. J. & Kastan, M. B. Initiating cellular stress responses. *Cell* **118**, 9–17 (2004).
27. Cuadrado, M. *et al.* ATM regulates ATR chromatin loading in response to DNA double-strand breaks. *J. Exp. Med.* **203**, 297–303 (2006).
28. Shao, R.-G. *et al.* Replication-mediated DNA damage by camptothecin induces phosphorylation of RPA by DNA-dependent protein kinase and dissociates RPA:DNA-PK complexes. *EMBO J.* **18**, 1397–1406 (1999).
29. Sakasai, R. *et al.* Differential involvement of phosphatidylinositol 3-kinase-related protein kinases in hyperphosphorylation of replication protein A2 in response to replication-mediated DNA double-strand breaks. *Genes Cells* **11**, 237–246 (2006).
30. Lukas, C., Falck, J., Bartkova, J., Bartek, J. & Lukas, J. Distinct spatiotemporal dynamics of mammalian checkpoint regulators induced by DNA damage. *Nature Cell Biol.* **5**, 255–260 (2003).
31. Bekker-Jensen, S. *et al.* Spatial organization of the mammalian genome surveillance machinery in response to DNA strand breaks. *J. Cell Biol.* **173**, 195–206 (2006).
32. Ball, H. L., Myers, J. S. & Cortez, D. ATRIP binding to RPA-ssDNA promotes ATR-ATRIP localization but is dispensable for Chk1 phosphorylation. *Mol. Biol. Cell* **16**, 2372–2381 (2005).
33. Adams, K. E., Medhurst, A. L., Dart, D. A. & Lakin, N. D. Recruitment of ATR to sites of ionising radiation-induced DNA damage requires ATM and components of the MRN protein complex. *Oncogene* **25**, 3894–3904 (2006).
34. Myers, J. S. & Cortez, D. Rapid activation of ATR by ionizing radiation requires ATM and Mre11. *J. Biol. Chem.* **281**, 9346–9350 (2006).
35. Paull, T. T. & Gellert, M. The 3' to 5' exonuclease activity of Mre11 facilitates repair of DNA double-strand breaks. *Mol. Cell* **1**, 969–979 (1998).
36. Assenmacher, N. & Hopfner, K. P. MRE11/RAD50/NBS1: complex activities. *Chromosoma* **113**, 157–166 (2004).
37. Larson, E. D., Cummings, W. J., Bednarski, D. W. & Maizels, N. MRE11/RAD50 cleaves DNA in the AID/UNG-Dependent Pathway of immunoglobulin gene diversification. *Mol. Cell* **20**, 367–375 (2005).
38. Trujillo, K. M., Yuan, S.-S. F., Lee, E.-Y. P. & Sung, P. Nuclease activities in a complex of human recombination and DNA repair factors Rad50, Mre11, and p95. *J. Biol. Chem.* **273**, 21447–21450 (1998).
39. Pierce, A. J., Hu, P., Han, M., Ellis, N. & Jasin, M. Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells. *Genes Dev.* **15**, 3237–3242 (2001).
40. Schaffer, A. A. *et al.* Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* **29**, 2994–3005 (2001).
41. Deng, C., Brown, J. A., You, D. & Brown, J. M. Multiple endonucleases function to repair covalent topoisomerase I complexes in *Saccharomyces cerevisiae*. *Genetics* **170**, 591–600 (2005).
42. Lisby, M., Barlow, J. H., Burgess, R. C. & Rothstein, R. Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. *Cell* **118**, 699–713 (2004).
43. Li, S. *et al.* Functional link of BRCA1 and ataxia telangiectasia gene product in DNA damage response. *Nature* **13**, 210–215 (2000).
44. Baroni, E., Viscardi, V., Cartagena-Lirola, H., Lucchini, G. & Longhese, M. P. The functions of budding yeast Sae2 in the DNA damage response require Mec1- and Tel1-dependent phosphorylation. *Mol. Cell Biol.* **24**, 4151–4165 (2004).
45. Cartagena-Lirola, H., Guerini, I., Viscardi, V., Lucchini, G. & Longhese, M. P. Budding yeast Sae2 is an *in vivo* target of the Mec1 and Tel1 checkpoint kinases during meiosis. *Cell Cycle* **5**, 1549–1559 (2006).
46. Vilkkii, S. *et al.* Screening for microsatellite instability target genes in colorectal cancers. *J. Med. Genet.* **39**, 785–789 (2002).
47. Wu, G. & Lee, W. H. CtIP, a multivalent adaptor connecting transcriptional regulation, checkpoint control and tumor suppression. *Cell Cycle* **5**, 1592–1596 (2006).
48. Raderschall, E., Golub, E. I. & Haaf, T. Nuclear foci of mammalian recombination proteins are located at single-stranded DNA regions formed after DNA damage. *Proc. Natl Acad. Sci. USA* **96**, 1921–1926 (1999).
49. Wu-Baer, F., Lagazon, K., Yuan, W. & Baer, R. The BRCA1/BARD1 heterodimer assembles polyubiquitin chains through an unconventional linkage involving lysine residue K6 of ubiquitin. *J. Biol. Chem.* **278**, 34743–34746 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Paull and J. Falck for providing reagents, and P. Huertas, A. Meier, K. Dry, K. Miller and Y. Pommier for advice and critical reading of the manuscript. This study was supported by Cancer Research UK, the EU and a Swiss National Science Foundation fellowship for advanced researcher (A.A.S.). Research in the S.P.J. laboratory is made possible by core infrastructure funding from Cancer Research UK and the Wellcome Trust. C.L., J.L., M.M. and J.B. were supported by grants from the Danish Cancer Society, Danish National Research Foundation, EU (DNA Repair), and the John and Birthe Meyer Foundation and the Czech Ministry of Education. Research in the laboratory of R.B. is supported by the National Institute of Health and S.F. is supported by a fellowship from the New York State Breast Cancer Research Fund.

Author Contributions S.F. and R.B. generated CtIP cDNA, CtIP antibodies and recombinant CtIP protein. C.L., J.L., M.M. and J.B. generated the cell lines with GFP-tagged proteins, conceived, performed and evaluated the real-time imaging experiments, and performed the homologous recombination measurements. All other experiments were conceived by A.A.S. and S.P.J., and were performed by A.A.S. with the help of J.C. A.A.S. and S.P.J. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.P.J. (s.jackson@gurdon.cam.ac.uk).

METHODS

Immunofluorescence microscopy. U2OS cells were transfected with siRNA and after 3 days were treated with 1 μ M camptothecin (CPT) for 1 h. After pre-extraction for 5 min on ice (25 mM Hepes 7.4, 50 mM NaCl, 1 mM EDTA, 3 mM MgCl₂, 300 mM sucrose and 0.5% TritonX-100), cells were fixed with 4% formaldehyde (w/v) in PBS for 12 min. Coverslips were washed with PBS and then co-immunostained with primary antibodies against γ H2AX (rabbit polyclonal, Cell Signalling Technology) either in combination with an antibody against RPA2 (mouse monoclonal, Lab Vision) or BrdU (mouse monoclonal, GE Healthcare). Appropriate Alexa Fluor-488 (green) and -594 (red) conjugated secondary antibodies (1:1,000) were purchased from Molecular Probes. Slides were viewed with a Bio-Rad confocal laser microscope by sequential scanning of the two emission channels used.

Immunoblotting. Cell extracts were prepared in Laemmli buffer (4% SDS, 20% glycerol, 120 mM Tris-HCl pH 6.8), proteins were resolved by SDS-PAGE and transferred to nitrocellulose. Immunoblots were performed by using the appropriate antibodies (see Supplementary Methods).

Immunoprecipitation. If not specified otherwise, cells were lysed in RIPA buffer (50 mM Tris-HCl, pH 7.4, 1% NP-40, 0.25% Na-deoxycholate, 150 mM NaCl, 1 mM EDTA and 0.1% SDS), supplemented with phosphatase inhibitor cocktails 1 and 2 (Sigma) and with complete protease inhibitor cocktail (Roche). Clarified extracts were pre-cleared with protein A or protein G beads (Sigma) for 1 h at 4 °C. Immunoprecipitating antibodies were added to the pre-cleared supernatant and incubated for 2 h at 4 °C. After a 1 h incubation with protein A or protein G beads, precipitated immunocomplexes were washed four times in lysis buffer (containing 0.5% NP-40, without SDS), boiled in SDS sample buffer and loaded on an SDS-polyacrylamide gel. Proteins were analysed by immunoblotting, as described above.

GST pull-downs. GST-CtIP fragments were constructed by cloning polymerase chain reaction products into pGEX-4T1 (Amersham). GST fusion plasmids were grown in BL21 RIL (Codon+) *Escherichia coli* (Stratagene) and protein was expressed by incubation for 4 h at 30 °C after the addition of 100 μ M IPTG. Proteins were purified from soluble extracts with glutathione Sepharose 4 fast flow beads (Amersham). GST fusion proteins bound to glutathione beads were mixed with 1 mg of HeLa nuclear extract and incubated for 1 h at 4 °C in 1 ml of TEN100 buffer (20 mM Tris-HCl, pH 7.4, 0.1 mM EDTA and 100 mM NaCl). Beads were then washed four times with NTEN300 buffer (0.5% NP-40, 0.1 mM EDTA, 20 mM Tris-HCl, pH 7.4, 300 mM NaCl), complexes were boiled in SDS sample buffer and analysed by SDS-PAGE followed by either Coomassie staining or immunoblotting.

In vitro nuclease assay. The PhiX174 circular single-stranded virion DNA substrate (5,386 nucleotides, New England Biolabs), was mixed with MRE11-RAD50 together with BSA or CtIP in reaction buffer (30 mM potassium-MOPS, pH 7.2, 1 mM dithiothreitol, 1 mM ATP, 25 mM KCl and 5 mM of either MgCl₂ or MnCl₂). After incubation for 3 h at 37 °C, nuclease reactions were terminated by adding 1/10 volume of stop solution (3% SDS, 50 mM EDTA) and samples were run in a 0.8% agarose gel (1 \times TAE) for 90 min at 100 mA. DNA species were stained with SYBR Gold Nucleic Acid Gel Stain (Molecular Probes) for 20 min and visualized with a fluoroimager (FLA-5000, Fujifilm).

Homologous recombination assay. A U2OS clone with the integrated homologous recombination reporter DR-GFP was generated as described previously³⁹. Two days after transfection with siRNA, U2OS DR-GFP cells were co-transfected with an I-SceI expression vector (pCBA-I-SceI) together with a vector expressing monomeric red fluorescent protein (pCS2-mRFP). The latter plasmid was added in 1:3 ratio to mark the I-SceI-positive cells. Cells were harvested two days after I-SceI transfection and subjected to flow cytometric analysis to examine recombination induced by double-strand breakage. Only mRFP-positive subpopulations of cells were analysed for homologous recombination efficiency to circumvent possible differences in transfection efficiencies. Fluorescence-activated cell sorting data were analysed with CellQuest software to reveal the percentage of GFP-positive cells relative to the number of transfected cells (mRFP positive). The data were then related to a control siRNA treatment in each individual experiment. The dividing line between GFP (homologous recombination) positive and negative cells was set to 1% of the background level of GFP-positive cells in every internal control (not transfected with I-SceI). This gate was then applied to the mRFP/I-SceI positive counterparts to determine homologous recombination efficiency (see Supplementary Fig. 5b).

Crystal structure of a catalytic intermediate of the maltose transporter

Michael L. Oldham¹, Dheeraj Khare¹, Florante A. Quiocho³, Amy L. Davidson² & Jue Chen¹

The maltose uptake system of *Escherichia coli* is a well-characterized member of the ATP-binding cassette transporter superfamily. Here we present the 2.8-Å crystal structure of the intact maltose transporter in complex with the maltose-binding protein, maltose and ATP. This structure, stabilized by a mutation that prevents ATP hydrolysis, captures the ATP-binding cassette dimer in a closed, ATP-bound conformation. Maltose is occluded within a solvent-filled cavity at the interface of the two transmembrane subunits, about halfway into the lipid bilayer. The binding protein docks onto the entrance of the cavity in an open conformation and serves as a cap to ensure unidirectional translocation of the sugar molecule. These results provide direct evidence for a concerted mechanism of transport in which solute is transferred from the binding protein to the transmembrane subunits when the cassette dimer closes to hydrolyse ATP.

ATP-binding cassette (ABC) transporters are ubiquitous membrane protein complexes that use the energy generated from ATP hydrolysis to transport solutes across membranes¹. Most ABC transporters contain two transmembrane domains or subunits (TMDs) and two cytosolic nucleotide-binding domains or subunits (NBDs), also known as ATP-binding cassettes. Prokaryotic ABC transporters involved in uptake contain a fifth component, a periplasmic or cell-surface-associated binding protein that binds specific solutes with high affinity². Whereas the sequences of the TMDs are diverse, those of the NBDs are highly conserved, with more than 30% identity between different ABC proteins³. In addition to the classical nucleotide-binding Walker A and B motifs⁴, each NBD also contains a conserved LSGGQ signature motif that is diagnostic of ABC ATPases and a Q loop that mediates the interactions between the TMD and the NBD. Crystal structures of four ABC transporters^{5–8} suggest at least two major states: an inward-facing conformation with the substrate translocation pathway open to the cytoplasm, and an outward-facing conformation with the translocation pathway facing the opposite side of the membrane.

One of the best functionally characterized ABC transporters is the maltose uptake system of *Escherichia coli*, which transports malto-oligosaccharides up to seven glucose units long⁹. Mapping of the genes encoding the maltose transporter in the 1970s^{10,11} showed that the system is composed of a periplasmic maltose-binding protein (MBP), two integral membrane proteins, MalF and MalG, and two copies of the cytoplasmic ATP-binding cassette MalK¹². Mutation studies have provided a wealth of information about transporter function², and transport activity has been reconstituted in proteoliposome vesicles, allowing biochemical characterization^{13–15}. Crystal structures of the closed and open states of MBP reveal a cleft between two lobes in which the oligosaccharide substrate binds and is completely engulfed by rotation of the lobes^{16–18}. While delivering the substrate to the transmembrane subunits, the binding protein also stimulates ATP hydrolysis¹⁵, ensuring that substrate is transported during each cycle of ATP hydrolysis. The observation that MBP is trapped in a complex with MalFGK₂ in the presence of a transition-state analogue (Mg·ADP·vanadate) suggests that binding of MBP induces a conformational change in the transporter and thereby triggers ATP hydrolysis¹⁹.

Previously, we reported the crystal structures of a MalK dimer in three different conformations^{20,21}. In the nucleotide-free forms, the amino-terminal NBDs are well separated and the dimer is maintained through contacts of the carboxy-terminal regulatory domains. We refer to this open conformation as the resting state. In the ATP-bound form, the NBDs are in close contact with two ATPs engulfed along the dimer interface and engaged in numerous contacts with residues from both monomers. In the ADP-bound form, the two NBDs are separated similarly to the nucleotide-free form, suggesting that ADP is insufficient to stabilize NBD closure and therefore that ATP hydrolysis may reset the transporter to the resting state. The conformational changes observed in the MalK dimer are presumably coupled with movements of the TMDs to facilitate the substrate's release from the binding protein and its subsequent translocation across the membrane.

We report here a 2.8-Å structure of the entire maltose transport complex, capturing a catalytic intermediate state. This structure, interpreted in the light of available genetic and biochemical data, shows that transfer of the substrate from MBP to MalFGK₂ is coupled to formation of the closed NBD dimer. Our structure, together with that of ModB₂C₂A (ref. 7), forms the basis of a model involving an alternating access mechanism in an ABC transport cycle.

Stabilizing the MBP–MalFGK₂ complex

To obtain a structure of the maltose transporter engaged with its binding protein, several transition-state analogues of ATP as well as mutant transporters were screened for crystallization. Crystals of MBP–MalFGK₂ were obtained from a mutant containing a glutamate-to-glutamine substitution (E159Q) located proximal to the γ-phosphate of ATP in the nucleotide-binding site. The corresponding mutation in BmrA²² and MJ0796 (ref. 23) abolishes ATP hydrolysis, trapping ATP in the nucleotide-binding site. Furthermore, in the presence of ATP, this substitution increases the affinity of the isolated NBDs for each other 10⁴-fold, permitting crystallization of the NBDs as a closed, ATP-bound dimer^{23,24}. In the intact maltose transport system, the E159Q mutant is also defective in hydrolysis, and ATP binding traps a complex of MBP and MalFGK₂ (A.L.D., unpublished observations).

¹Department of Biological Sciences, ²Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, USA. ³Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA.

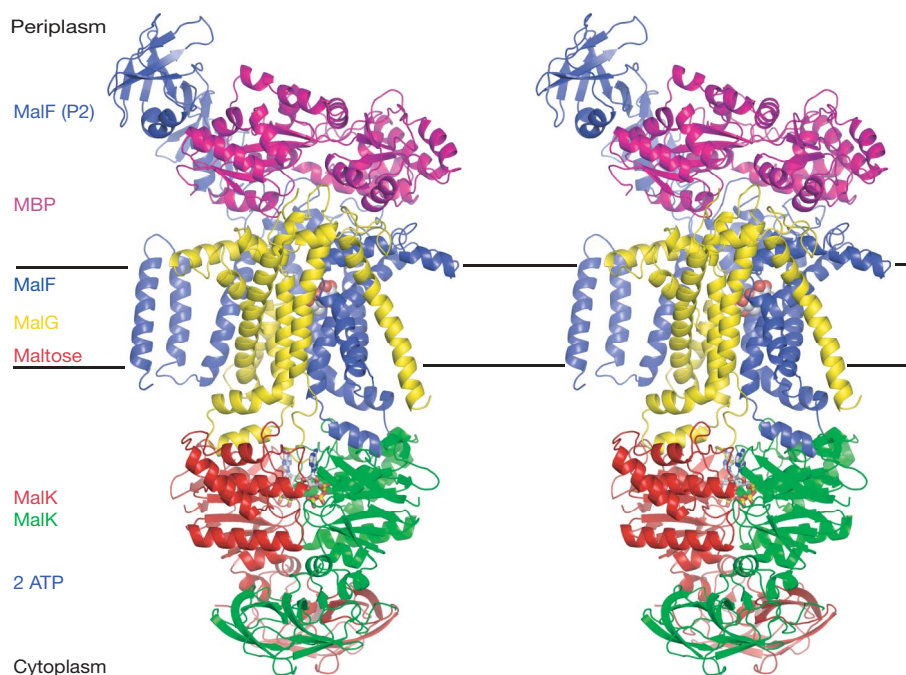


Figure 1 | Structure of the maltose transporter in a catalytic intermediate conformation. Stereo view of the complex in a ribbon diagram. Colour codes: magenta, MBP; blue, MalF; yellow, MalG; red and green, MalK dimer.

Overview of the structure

Crystals of MalFGK₂(E159Q) in complex with MBP, maltose and ATP diffracted to 2.8 Å. The structure was determined by molecular replacement²⁵, with MalK and MBP as search models (Supplementary Table 1). One MalK dimer and one MBP molecule were found by a rotation and translation search, and the resulting electron-density map showed clear densities for the transmembrane helices (Supplementary Fig. 1). Cycles of model building and refinement yielded a model containing 1884 residues ($R_{\text{work}} = 24.2\%$ and $R_{\text{free}} = 27.1\%$), with no clear electron density for the remaining 38 residues located in periplasmic loops or the termini of different subunits. The sequence assignments were confirmed by the 39 Se sites revealed in the difference Fourier density map from selenomethionine-substituted MalFGK₂ in complex with native MBP (Supplementary Fig. 2).

The five subunits of the maltose transporter assemble into a tight complex of about $165 \times 75 \times 75 \text{ Å}^3$ (Fig. 1). At the cytoplasmic side, the two MalK subunits form a closed dimer with two ATP molecules bound at the interface. The structures of the two transmembrane (TM) subunits, MalF and MalG, are composed of eight and six TM helices, respectively. The C-terminal segment of MalG is inserted partway into the MalK dimer. The maltose is bound at the interface of MalF and MalG about halfway into the lipid bilayer. At the periplasmic surface, MBP docks onto MalF and MalG in a ligand-free open form, with both lobes making extensive interactions with the TM subunits. Two notable interactions include the insertion of the MalG periplasmic loop (P3) into the MBP sugar-binding cleft and contacts by mostly the N-lobe of MBP with a large periplasmic loop (P2) of MalF, which folds into an Ig-like domain and extends about 30 Å away from the membrane surface.

An ATP-bound, closed MalK dimer

The structure of MalK₂(E159Q) in the complex is essentially identical to that of the isolated wild-type MalK dimer in the ATP-bound, closed form²⁰, as is evident from a root mean squared deviation of 0.6 Å for the C α positions between 740 residues of the dimer. Except for regions involved in crystal packing, the only significant difference between the isolated and membrane-bound MalK lies in the Q loop (residues 79–90), where MalK contacts the TM subunits. In the intact

Maltose and ATP molecules are shown in CPK and ball-and-stick models, respectively (red, O atom; grey, C atom).

transporter the Q loop is well ordered and contains a β -strand. The flexible nature of the Q loop observed in isolated MalK structures^{20,21} is probably due to the absence of the interacting transmembrane subunits. Two ATP molecules are bound along the dimer interface, contacting residues in the Walker A and B motifs from one monomer and the LSGGQ motif of the opposite monomer (Fig. 2). This closed architecture of the NBD dimer is also observed in the crystal structure of the bacterial exporter Sav1866 (ref. 8; Supplementary Fig. 3). The observation that both MalFGK₂ and Sav1866 adopt an outward-facing conformation in the ATP-bound state suggests that importers and exporters share a common mechanism of coupling ATP hydrolysis to substrate translocation.

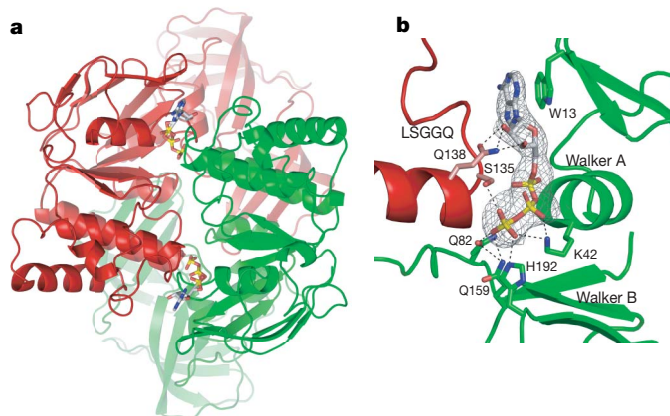


Figure 2 | Ribbon diagram of the MalK subunits with bound ATP. a, The MalK dimer viewed down the local twofold axis. The two subunits are coloured in red and green. The ATP is represented in ball-and-stick model. b, The ATP-binding site, showing that residues from both MalK subunits are making contact with the ATP. Interacting residues and associated hydrogen bonds (black dashed lines) from residues in the Walker A, Walker B and LSGGQ motifs, and from H192 and Q159, to ATP are indicated. A positive $F_o - F_c$ electron density (contoured at 1.5σ) obtained with ATP omitted in the structure factor calculation is also shown.

Symmetry of the TM helices

The topology of MalF and MalG agrees well with the predicted model based on genetic studies^{26–28} (Fig. 3a). Viewed from the periplasmic side, the TM helices of MalF and MalG assemble into two crescent-shaped structures with their concave surfaces facing each other and enclosing the maltose (Fig. 3b). TM helices 1–3 of MalF cross over the dimer interface to contact TM2–6 of MalG, and TM1 of MalG packs against the helical bundle formed by TM4–8 of MalF (Figs 1 and 3). This intertwining configuration, also seen in the structure of ModB₂C₂A (ref. 7), is quite different from the side-by-side arrangement in the structures of BtuCD^{6,29} and H1470/1 (ref. 5). MalFGK₂ is unique among these ABC transporters because its TMDs are assembled from a heterodimer instead of a homodimer. Moreover, the lengths of MalF and MalG are different (514 and 296 residues, respectively) and share only 13% sequence identity. Nevertheless, a pseudo-twofold symmetry is observed in the core region of the TM subunits, with TM3–8 of MalF related to TM1–6 of MalG by about 180° rotation along an axis perpendicular to the predicted location of the membrane bilayer (Fig. 3b).

The MalK–TMD interface

The MalK subunits are bound to MalF and MalG primarily through contacts with the 'EAA' motif (EAA-X(3)-G)^{30,31} of the TM subunits (Fig. 4 and Supplementary Fig. 4). The structures of the two EAA loops of MalF and MalG are very similar, each consisting of two short helices.

One of the two helices, previously named the coupling helix⁷, docks into a surface cleft on each MalK subunit (Fig. 4a). The MalK cleft consists of two helices from the helical subdomain²⁰, the helix following the Walker A motif and residues in the Q loop. In both MalF and MalG, the glutamate residue in the EAA loop is engaged in a salt-bridge interaction with an arginine residue (R47; Fig. 4b). This arginine is highly conserved in carbohydrate uptake systems I (ref. 32), where the EAA loop is also conserved. Replacement of the glutamate in the EAA loop in either MalF or MalG has little effect on maltose transport, but the double mutant is completely defective and results in cytosolic localization of MalK³³. Another critical contact is made by a residue on the coupling helix, namely M405 in MalF and L194 in MalG. The side chain of these residues inserts into a hydrophobic pocket formed by MalK residues A50, L52, A73, V77, M79 and F81 (Fig. 4b).

One of the unexpected features revealed by the MBP–MalFGK₂ structure is the insertion of the MalG C-terminal tail into the MalK dimer interface. Residues 290–296 of MalG pack along the Q loop (residues 83–89) of one MalK monomer, with the terminal carboxyl group of MalG making three hydrogen bonds with backbone atoms of the Q loop in the other monomer (residues S83, A85 and L86; Fig. 4c). Although the C-terminal tail of MalG is not required for closure of the isolated MalK dimer²⁰, the coupling between the two MalK monomers by MalG seems to contribute to the ordering of the Q loop and may be important in the formation of the catalytic intermediate conformation of the entire transporter.

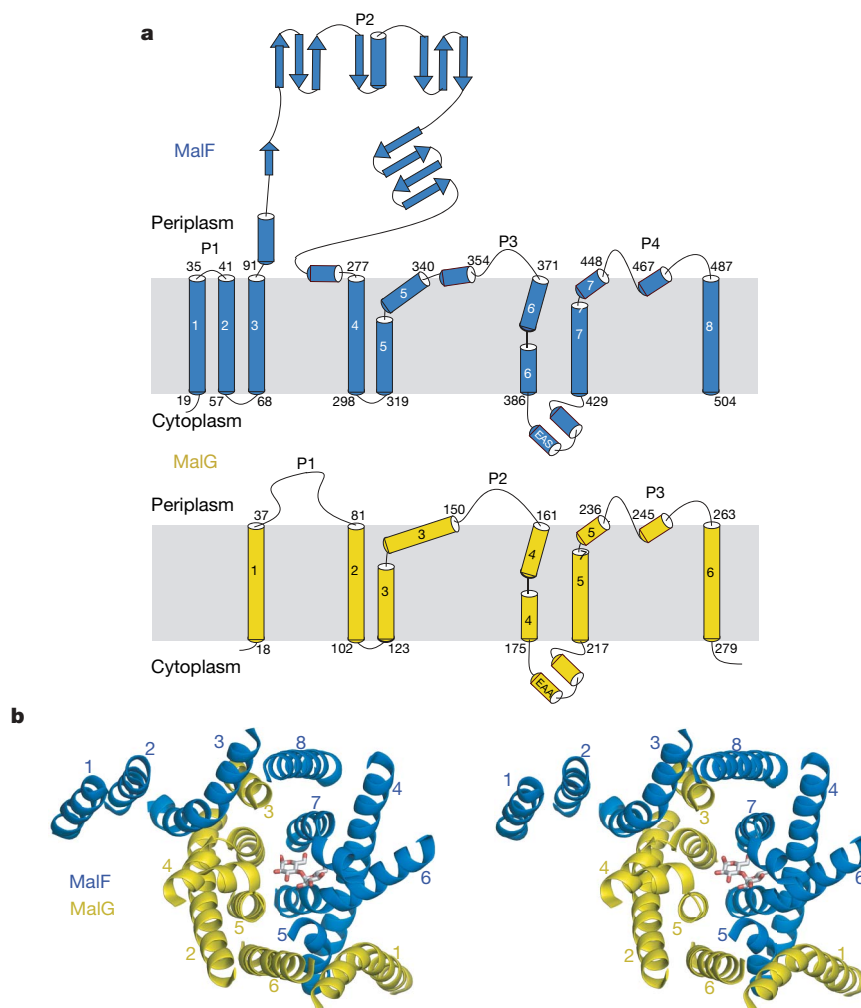


Figure 3 | Architecture of the transmembrane subunits. **a**, Diagram of the secondary structure of MalF (blue) and MalG (yellow). The first and last residues of each TM helix predicted to be buried by the lipid acyl chains are shown in the figure. **b**, Ribbon representation of the 14 TM helices viewed

along the membrane normal from the periplasmic side (stereo view). Loops are omitted for clarity. Helices 3–8 of MalF (blue) are related to helices 1–6 of MalG (yellow) by rotation about 180°. Maltose is shown in ball-and-stick model (red, O atom; grey, C atom).

The substrate translocation pathway

A large solvent-filled cavity, located at the interface of MBP, MalF and MalG, reaches halfway across the predicted membrane bilayer from the periplasmic surface (Fig. 5a, b). The volume of this cavity is about 6,500 Å³, sufficient to accommodate the largest oligosaccharide (maltoheptaose) that can be transported. This substrate translocation pathway is completely shielded from the membrane bilayer by TM helices of MalF and MalG and is further sealed by MBP at the periplasmic side (Supplementary Fig. 5). The translocation pathway is accessible through several narrow solvent channels connecting the TM cavity with the periplasm. However, the largest radius of these channels is about 5 Å, sufficient for water molecules to diffuse through but too small for maltose to escape back to the periplasm.

At the bottom of the cavity, clear electron density for maltose is observed (Fig. 5c). A maltose molecule was placed in an orientation that fits the density and makes complementary interactions with surrounding residues. The aromatic side chains of MalF Y383 and F436 make stacking interactions with the sugar rings. Hydrogen-bonding and van der Waals interactions with MalF Y325, S329, N376, L379, G380, S433, N437 and N440 further confer specificity and stabilize substrate binding. Out of the 10 residues that contact maltose, 6 had been identified through genetic studies^{34,35} to severely decrease or eliminate maltose transport (Supplementary Table 2). No residue from MalG makes contact with maltose. The chemical environment of this sugar-binding site is similar to that of MBP, with both aromatic residue stacking and hydrogen bonding contributing to sugar recognition and binding³⁶. Our studies provide the first structural evidence for the presence of a substrate-binding site inside the TM subunits of an ABC transporter, as previously suggested by genetic data³⁷. There is apparently only a single binding site, suggesting one maltose per transport cycle. The mode of interaction between the substrate and the residues in the cavity is markedly different from that of maltoporin, an outer-membrane channel³⁸, in which the extended hydrophobic path, the 'greasy slide', would allow a continuous flow of maltose down the concentration gradient. The two

very different translocation pathways seen in the maltose transport systems may reflect fundamental differences in the operation of channels versus transporters.

The open apo conformation of MBP in this intermediate form has significant mechanistic implications. The crystal structures of isolated MBP^{18,39} show that the open form has a lower affinity for maltose because it makes fewer contacts than the closed form with the sugar. Whereas in solution maltose shifts MBP from the open to the closed form^{17,40}, interactions with the TM subunits stabilize MBP in the open state, presumably to promote sugar release. The absence of maltose electron density in the binding site of MBP demonstrates that maltose has been translocated to the TM subunits.

Another notable feature at the MBP–TM interface relating to maltose transfer is the insertion of the MalG P3 loop into the MBP sugar-binding site. If one models a sugar molecule into the binding site based on the structures of substrate-bound, open forms of MBP¹⁸, even the smallest substrate, maltose, clashes with the P3 loop (Fig. 5d). Insertion of the P3 loop, the 'scoop', would therefore dislodge the sugar from the binding protein and ensure efficient transfer. Consistently, a mutant containing a 31-residue insertion in this loop was defective in transport but retained the ability to assemble a MalFGK₂ complex⁴¹. Insertion of periplasmic loops into the binding protein was also observed in the recent structure of BtuCD-F²⁹.

Discussion

The structure of the MBP–MalFGK₂(E159Q) complex represents an intermediate state in which the NBDs are closed to hydrolyse ATP and the maltose has already been translocated from MBP into the transmembrane subunits. It suggests that MBP stimulates ATP hydrolysis by triggering a conformational change that results in the closure of the NBDs. This structure shows several features of the transition-state conformation characterized by biochemical studies, including the strong interactions between MalFGK₂ and MBP¹⁹, the opened sugar-binding cleft of the MBP⁴², and the closed state of the NBD dimer⁴³. However, the presence of the E159Q mutation in both

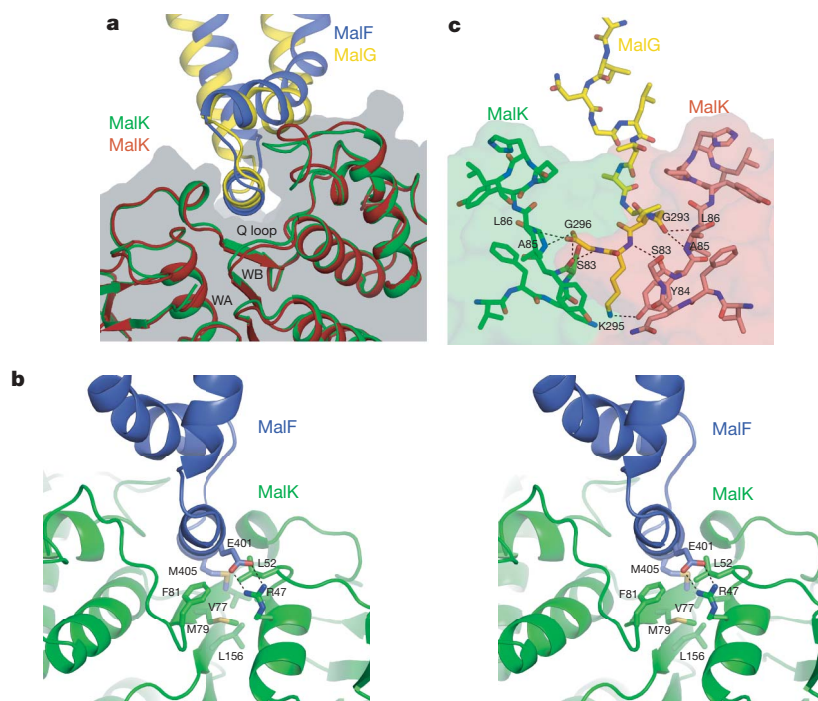


Figure 4 | The TMD–MalK interface. **a**, Docking of the EAA loops into a surface cleft of MalK. The EAA loops of MalF and MalG are compared by superposition of the two MalK subunits. The MalK dimer is also shown as a transparent surface model. WA, Walker A motif; WB, Walker B motif. **b**, Stereo view of the interactions between MalF and MalK. **c**, Insertion of the

MalG C-terminal tail into the MalK dimer interface. The two MalK subunits are represented as a transparent surface model except for the interacting Q loops, which are shown in stick model. Hydrogen bonds and salt bridges are indicated by black dashed lines.

MalK subunits prevents hydrolysis and generates symmetry between nucleotide-binding sites that may not exist in the catalytic transition state^{44–46}.

Comparison of the structure of MalFGK₂ with ModB₂C₂A (ref. 7), an archaeal molybdate importer, reveals a striking similarity in the TM region despite a lack of significant sequence similarity (Fig. 6a), suggesting that their mechanisms of translocation are very similar. The proposed gating region in ModB is positioned around the substrate-binding site in MalF (Fig. 6a). In addition, the structure of the MalFGK₂ EAA loop is essentially identical to that of ModB (Fig. 6b). Superposition of the NBDs as they occur in the MalFGK₂ and ModB₂C₂A structures shows that the coupling helices rotate about 20° inside a cleft on the NBD surface, a rotation similar to the motion of a ball-and-socket joint (Fig. 6c). Structural comparison of the two complete transporters shows an 11-Å difference in the distance between the two coupling helices within a transporter that correlates with an interchange between two conformational states.

The structures of MBP–MalFGK₂ and ModB₂C₂A (ref. 7) support a model involving an alternating-access mechanism⁴⁷ in an ABC transport cycle (Fig. 7). Whereas our structure probably represents the catalytic transition-state conformation with the translocation pathway facing outwards, the structure of ModB₂C₂A represents

the transporter in a resting-state conformation with the translocation pathway facing inwards (Fig. 7). In the resting state, the NBD dimer interface is open and the two coupling helices are farther apart. Interactions with the closed, substrate-loaded binding protein promote the transition from this resting state to the transition state, in which the NBD dimer is closed for hydrolysis and the binding protein is stabilized in an open conformation with decreased affinity for the substrate. After ATP hydrolysis, the binding protein disengages from the transporter, the NBD dimer opens and the transporter reverts to the resting state concomitantly with exposure of the substrate to the cytoplasm. Reorientation of the TM helices may also abolish the TM substrate-binding site, which is formed by residues from three different TM helices in our structure (Fig. 5), thus releasing the substrate. A key concept of this model is that the transition from the resting to transition state is driven by the engagement of the substrate-loaded binding protein instead of the binding of ATP to the Walker A and B motifs. The resting-state structure of ModB₂C₂A was obtained with the binding protein in the absence of nucleotides, but *in vivo* the nucleotide-binding sites are probably saturated with ATP, because the intracellular ATP concentration⁴⁸ is ten times the *K_m* of most ABC transporters^{49,50}. Whereas isolated NBDs form a closed dimer on binding ATP, the NBD dimer interface would be held open by the TMDs in the resting state of an intact transporter,

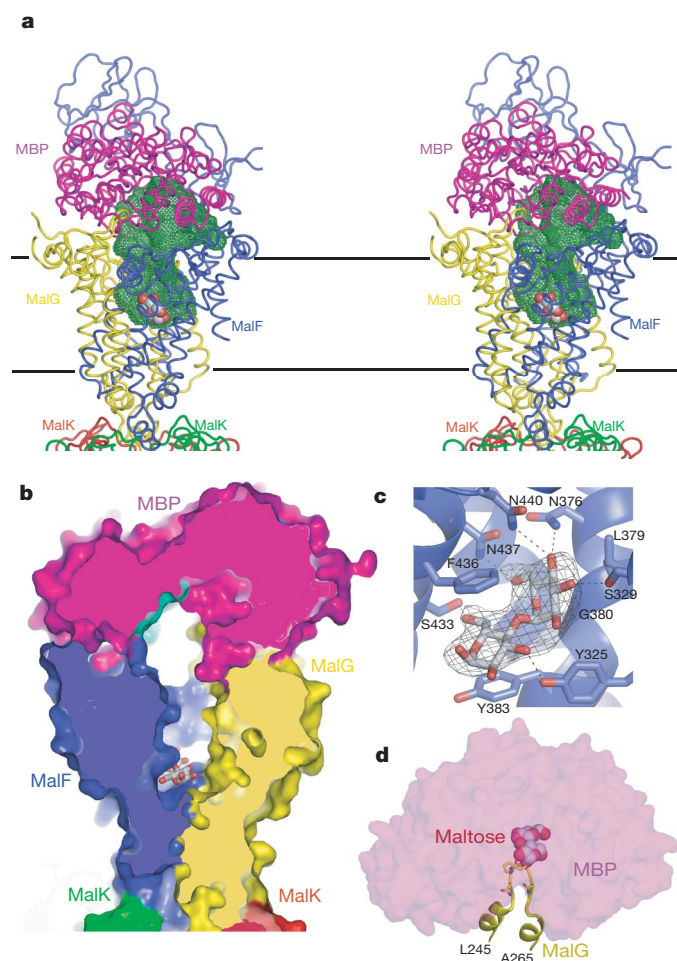


Figure 5 | Transfer of the maltose from MBP to the TM binding site.

a, Stereo view of the occluded maltose in a large cavity (green mesh). **b**, Slab view of the cavity, with the proteins shown in space-filling models. Residues of MBP that form the substrate-binding site in the open conformation are shown in cyan. **c**, A close-up view of the TM maltose-binding site with hydrogen bonds indicated. Also shown is the $F_o - F_c$ electron density (2.5σ) calculated with maltose omitted. **d**, Insertion of the MalG scoop loop into the substrate-binding site of MBP. A maltose molecule is modelled into the binding site on the basis of the crystal structure of open maltose-bound MBP (PDB accession number 1JW5)¹⁸.

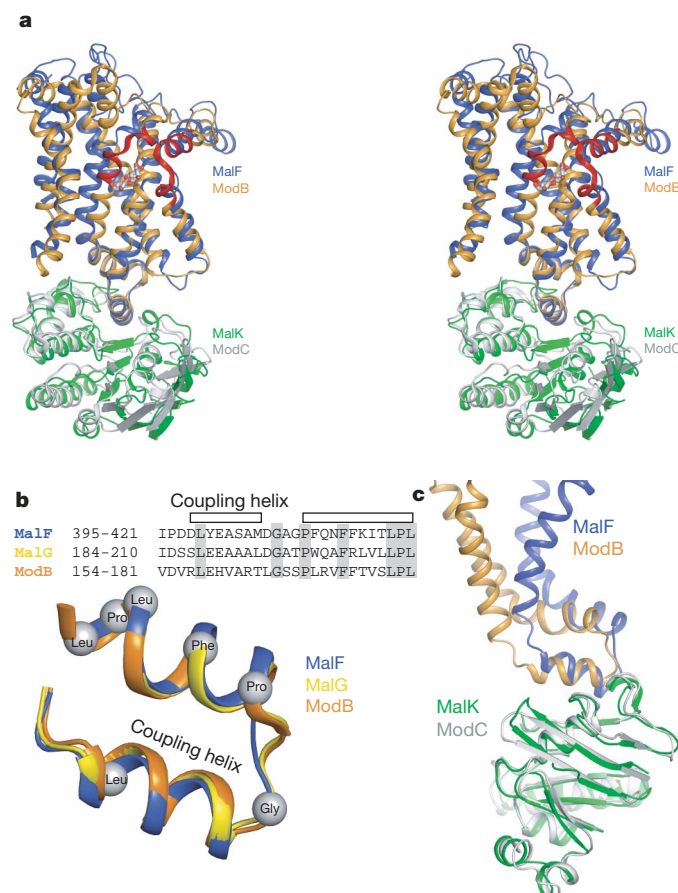


Figure 6 | Comparison of the structures of two binding-protein-dependent ABC uptake systems.

a, Similar topology of the TMDs. TM helices of MalF (blue) and ModB (orange) (PDB accession number 2ONK)⁷ and their interacting NBDs (green, MalK; grey, ModC) are superimposed, aligned on the basis of the EAA loops. The TM1 and TM2 helices and the P2 loop of MalF are omitted for clarity. The proposed gating regions⁷ in ModB are highlighted in red. Maltose is shown in stick model. **b**, Structure-based sequence alignment and superposition of the EAA loops of MalF (blue), MalG (yellow) and ModB (orange). Conserved residues are highlighted in grey. **c**, Comparison of the cytoplasmic helices of MalF and ModB, with the interacting NBDs superimposed.

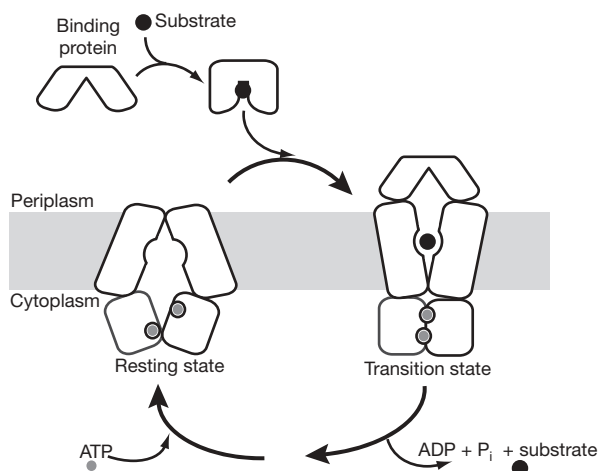


Figure 7 | A model for the ABC uptake system. The resting state conformation is modelled on the basis of the crystal structures of ModB₂C₂A (ref. 7) and H11470/1 (ref. 5). The transition state is based on the maltose transporter structure determined in this study.

even with ATP bound to the Walker A motifs, awaiting substrate to transport. Efficient stimulation of hydrolysis by MBP requires maltose¹⁵, suggesting that only the closed, substrate-loaded binding protein, not the open apo form, initiates the formation of the catalytic transition state. It is generally accepted that in exporters, substrate binding modulates the ATPase activity. We suggest that importers function by a similar mechanism, except that the substrate acts through its interactions with the binding protein.

METHODS SUMMARY

MBP and MalFGK₂ were expressed separately in *E. coli* as described previously¹⁹. Membrane-solubilized MalFGK₂ was purified by affinity and gel-filtration chromatography in the presence of n-dodecyl-β-D-maltoside and n-undecyl-β-D-maltoside, respectively. MBP was purified by ion-exchange and gel-filtration chromatography in the absence of detergent. The MBP–MalFGK₂ complex was prepared by mixing excess MBP with MalFGK₂ in the presence of ATP, EDTA and maltose, and purified by gel-filtration chromatography. Diffraction data were collected at the beamline 23ID at the Advanced Photon Source (APS). The structure was solved by molecular replacement²⁵ from structures of the closed MalK (1Q12) and the open MBP (1OMP). Figures were prepared with the program PyMOL (<http://www.pymol.org>).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 August; accepted 17 September 2007.

- Higgins, C. F. ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol.* **8**, 67–113 (1992).
- Boos, W. & Lucht, J. M. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* (eds Neidhardt, F. C. et al.) 1175–1209 (ASM Press, Washington DC, 1996).
- Holland, I. B. & Blight, M. A. ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. *J. Mol. Biol.* **293**, 381–399 (1999).
- Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945–951 (1982).
- Pinkett, H. W., Lee, A. T., Lum, P., Locher, K. P. & Rees, D. C. An inward-facing conformation of a putative metal-chelate-type ABC transporter. *Science* **315**, 373–377 (2007).
- Locher, K. P., Lee, A. T. & Rees, D. C. The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* **296**, 1091–1098 (2002).
- Hollenstein, K., Frei, D. C. & Locher, K. P. Structure of an ABC transporter in complex with its binding protein. *Nature* **446**, 213–216 (2007).
- Dawson, R. J. & Locher, K. P. Structure of a bacterial multidrug ABC transporter. *Nature* **443**, 180–185 (2006).

- Ferenci, T. The recognition of maltodextrins by *Escherichia coli*. *Eur. J. Biochem.* **108**, 613–636 (1980).
- Raibaud, O., Roa, M., Braun-Breton, C. & Schwartz, M. Structure of the *malB* region in *Escherichia coli* K-12. I. Genetic map of the *malK-lamB* operon. *Mol. Gen. Genet.* **174**, 241–248 (1979).
- Silhavy, T. J. et al. Structure of the *malB* region in *Escherichia coli* K12. II. Genetic map of the *malE,F,G* operon. *Mol. Gen. Genet.* **174**, 249–259 (1979).
- Davidson, A. L. & Nikaido, H. Purification and characterization of the membrane-associated components of the maltose transport system from *Escherichia coli*. *J. Biol. Chem.* **266**, 8946–8951 (1991).
- Davidson, A. L. & Nikaido, H. Overproduction, solubilization, and reconstitution of the maltose transport system from *Escherichia coli*. *J. Biol. Chem.* **265**, 4254–4260 (1990).
- Landmesser, H. et al. Large-scale purification, dissociation and functional reassembly of the maltose ATP-binding cassette transporter (MalFGK₂) of *Salmonella typhimurium*. *Biochim. Biophys. Acta* **1565**, 64–72 (2002).
- Davidson, A. L., Shuman, H. A. & Nikaido, H. Mechanism of maltose transport in *Escherichia coli*: Transmembrane signalling by periplasmic binding proteins. *Proc. Natl Acad. Sci. USA* **89**, 2360–2364 (1992).
- Sharff, A. J., Rodseth, L. E., Spurlino, J. E. & Quijcho, F. A. Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry* **31**, 10657–10663 (1992).
- Quijcho, F. A., Spurlino, J. C. & Rodseth, L. E. Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure* **5**, 997–1015 (1997).
- Duan, X. & Quijcho, F. A. Structural evidence for the dominant role of nonpolar interactions in the binding of a transport/chemosensory receptor to its highly polar ligands. *Biochemistry* **41**, 706–712 (2002).
- Chen, J., Sharma, S., Quijcho, F. A. & Davidson, A. L. Trapping the transition state of an ATP-binding-cassette transporter: Evidence for a concerted mechanism of maltose transport. *Proc. Natl Acad. Sci. USA* **98**, 1525–1530 (2001).
- Chen, J., Lu, G., Lin, J., Davidson, A. L. & Quijcho, F. A. A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Mol. Cell* **12**, 651–661 (2003).
- Lu, G., Westbrook, J. M., Davidson, A. L. & Chen, J. ATP hydrolysis is required to reset the ATP-binding cassette dimer into the resting-state conformation. *Proc. Natl Acad. Sci. USA* **102**, 17969–17974 (2005).
- Orelle, C., Dalmás, O., Gros, P., Di Pietro, A. & Jault, J. M. The conserved glutamate residue adjacent to the Walker-B motif is the catalytic base for ATP hydrolysis in the ATP-binding cassette transporter BmrA. *J. Biol. Chem.* **278**, 47002–47008 (2003).
- Moody, J. E., Millen, L., Binns, D., Hunt, J. F. & Thomas, P. J. Cooperative, ATP-dependent association of the nucleotide binding cassettes during the catalytic cycle of ATP-binding cassette transporters. *J. Biol. Chem.* **277**, 21111–21114 (2002).
- Smith, P. C. et al. ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol. Cell* **10**, 139–149 (2002).
- Rossmann, M. G. The molecular replacement method. *Acta Crystallogr. A* **46**, 73–82 (1990).
- Covitz, K.-M. Y., Panagiotidis, C. H., Reyes, M., Treptow, N. A. & Shuman, H. A. Mutations that alter the transmembrane signalling pathway in an ATP binding cassette (ABC) transporter. *EMBO J.* **13**, 1752–1759 (1994).
- Dassa, E. & Muir, S. Membrane topology of MalG, an inner membrane protein from the maltose transport system of *Escherichia coli*. *Mol. Microbiol.* **7**, 29–38 (1993).
- Froshauer, S., Green, G. N., Boyd, D., McGovern, K. & Beckwith, J. Genetic analysis of the membrane insertion and topology of MalF, a cytoplasmic membrane protein of *Escherichia coli*. *J. Mol. Biol.* **200**, 501–511 (1988).
- Hvorup, R. N. et al. Asymmetry in the structure of the ABC transporter binding protein complex BtuCD-BtuF. *Science* **317**, 1387–1390 (2007).
- Dassa, E. & Hofnung, M. Sequence of gene *malG* in *E. coli* K12: homologies between integral membrane components from binding protein-dependent transport systems. *EMBO J.* **4**, 2287–2293 (1985).
- Saurin, W., Koster, W. & Dassa, E. Bacterial binding protein-dependent permeases: characterization of distinctive signatures for functionally related integral cytoplasmic membrane proteins. *Mol. Microbiol.* **12**, 993–1004 (1994).
- Busch, W. & Saier, M. H. Jr. The transporter classification (TC) system, 2002. *Crit. Rev. Biochem. Mol. Biol.* **37**, 287–337 (2002).
- Mourez, M., Hofnung, M. & Dassa, E. Subunit interactions in ABC transporters: a conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits. *EMBO J.* **16**, 3066–3077 (1997).
- Ehrle, R., Pick, C., Ulrich, R., Hofmann, E. & Ehrmann, M. Characterization of transmembrane domains 6, 7, and 8 of MalF by mutational analysis. *J. Bacteriol.* **178**, 2255–2262 (1996).
- Steinke, A., Grau, S., Davidson, A., Hofmann, E. & Ehrmann, M. Characterization of transmembrane segments 3, 4, and 5 of MalF by mutational analysis. *J. Bacteriol.* **183**, 375–381 (2001).
- Quijcho, F. A. Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. *Annu. Rev. Biochem.* **55**, 287–315 (1986).

37. Shuman, H. A. Active transport of maltose in *Escherichia coli* K-12: role of the periplasmic maltose binding protein and evidence for a substrate recognition site in the cytoplasmic membrane. *J. Biol. Chem.* **257**, 5455–5461 (1982).
38. Schirmer, T., Keller, T. A., Wang, Y. F. & Rosenbusch, J. P. Structural basis for sugar translocation through maltoporin channels at 3.1 Å resolution. *Science* **267**, 512–514 (1995).
39. Quijcho, F. A. & Ledvina, P. S. Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol. Microbiol.* **20**, 17–25 (1996).
40. Miller, D. M., Olson, J. S., Pflugrath, J. W. & Quijcho, F. A. Rates of ligand binding to periplasmic proteins involved in bacterial transport and chemotaxis. *J. Biol. Chem.* **258**, 13665–13672 (1983).
41. Nelson, B. D. & Traxler, B. Exploring the role of integral membrane proteins in ATP-binding cassette transporters: analysis of a collection of MalG insertion mutants. *J. Bacteriol.* **180**, 2507–2514 (1998).
42. Austermuhle, M. I., Hall, J. A., Klug, C. S. & Davidson, A. L. Maltose-binding protein is open in the catalytic transition state for ATP hydrolysis during maltose transport. *J. Biol. Chem.* **279**, 28243–28250 (2004).
43. Fetsch, E. E. & Davidson, A. L. Vanadate-catalyzed photocleavage of the signature motif of an ATP-binding cassette (ABC) transporter. *Proc. Natl Acad. Sci. USA* **99**, 9685–9690 (2002).
44. Sharma, S. & Davidson, A. L. Vanadate-induced trapping of nucleotide by the purified maltose transport complex requires ATP hydrolysis. *J. Bacteriol.* **182**, 6570–6576 (2000).
45. Urbatsch, I. L., Sankaran, B., Bhagat, S. & Senior, A. E. Both P-glycoprotein nucleotide-binding sites are catalytically active. *J. Biol. Chem.* **270**, 26956–26962 (1995).
46. Urbatsch, I. L., Sankaran, B., Weber, J. & Senior, A. E. P-glycoprotein is stably inhibited by vanadate-induced trapping of nucleotide at a single catalytic site. *J. Biol. Chem.* **270**, 19383–19390 (1995).
47. Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969–970 (1966).
48. Neuhaud, J. & Nygaard, P. in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (eds Ingraham, J. L. & Neidhardt, F. C.) 445–473 (ASM Press, Washington DC, 1987).
49. Davidson, A. L., Laghaeian, S. S. & Mannering, D. E. The maltose transport system of *Escherichia coli* displays positive cooperativity in ATP hydrolysis. *J. Biol. Chem.* **271**, 4858–4863 (1996).
50. Sun, H. & Nathans, J. Mechanistic studies of ABCR, the ABC transporter in photoreceptor outer segments responsible for autosomal recessive Stargardt disease. *J. Bioenerg. Biomembr.* **33**, 523–530 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the beamline staff at the Advanced Photon Source beamline 23-ID for assistance with data collection, and R. MacKinnon, H. Shuman, D. Yernool and C. Orelle for discussions. This work was supported by NIH grants (J.C., A.L.D. and F.A.Q.), the Welch Foundation (F.A.Q.), the Pew Scholar Program (J.C.) and a postdoctoral fellowship from American Heart Association (M.L.O.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.C. (chenjue@purdue.edu).

METHODS

Expression and purification of MalFGK₂. Three antibiotic-resistance-compatible plasmids carrying *malF/malG* (*amp^R*), *malK*(E159Q) (*cam^R*) and *lacIq* (*spec^R*) were transformed into the HN741 host strain, [*E. coli* K-12 *argH his tpsL1 malt(Con) malBA13 AatpBC ilv::Tn10/F' lac^R Tn5*], which contains chromosomal deletions of the genes of the maltose transport system and the F₀F₁ ATPase⁵¹. An engineered C-terminal His₆ tag (ASASHHHHHH) used previously in the structural determination of the MalK homodimer⁵² was also used in this study. Cells were grown at 37 °C to mid-exponential phase in Terrific Broth medium (Terrific Broth I; MP Biomedicals) and induced with 50 μM isopropyl β-D-thiogalactoside at 22 °C for 24 h. After harvesting by centrifugation (4,000g for 12 min at 4 °C), the cells were broken by two passes through a high-pressure homogenizer (EmulsiFlex-C5; Avestin). The membrane fraction was isolated by centrifugation (100,000g for 30 min at 4 °C); the resulting membrane pellets were resuspended in buffer containing 20 mM Tris-HCl pH 8, 100 mM NaCl, 10% glycerol and 5 mM MgCl₂, and diluted to 5 mg ml⁻¹ total protein content as assayed by Bradford dye. Membranes were solubilized for 4 h with 0.35% n-dodecyl-β-D-maltoside (DDM; Anatrace). Insoluble debris was removed by centrifugation (100,000g for 20 min at 4 °C); the supernatant was applied to cobalt-affinity resin (Clontech) equilibrated with buffer containing 10 mM Tris-HCl pH 8, 100 mM NaCl, 10% glycerol and 0.01% DDM. The column was washed with equilibration buffer containing 5 mM imidazole and eluted with 100 mM imidazole. The fractions containing the transporter were pooled and dialysed against buffer containing 10 mM Tris-HCl pH 8, 200 mM NaCl and 0.01% DDM. Selenomethionine-labelled MalFGK₂ was expressed using the methionine auxotroph *E. coli* strain B834 (Stratagene) in minimal medium supplemented with nutrients (SelenoMet Medium; AthenaES) and selenomethionine (Acros). The purification was performed identically to that of the native transporter except for the addition of 5 mM 2-mercaptoethanol.

Expression and purification of MBP. MBP was produced similarly to the published protocol⁵³. In brief, cells of the strain BAR1000 harbouring increased expression of MBP under the *mal* promoter were cultured overnight in Luria-Bertani medium supplemented with 4 g l⁻¹ maltose and were harvested by centrifugation (4,000g for 12 min at 4 °C). MBP was purified by osmotic shock. Cells were resuspended gently in buffer containing 30 mM Tris-HCl pH 8, 20% sucrose and 1 mM EDTA. After centrifugation (10,000g for 10 min at 4 °C), the cell pellet was resuspended in ice-cold 5 mM MgSO₄ to disrupt the outer membrane and release MBP to the solution. The resulting spheroplasts were removed by centrifugation and the resulting supernatant containing MBP was dialysed against 20 mM Tris-HCl pH 8. MBP was further purified by ion-exchange chromatography (Source 15Q; GE Healthcare) and gel-filtration chromatography (Superdex 200; GE Healthcare) through FPLC (AKTA; GE Healthcare) at 4 °C.

Complex formation and crystallization. MBP and MalFGK₂(E159Q) (1 mg ml⁻¹) were mixed at a 1.25:1 stoichiometry in the presence of 1 mM ATP, 1 mM MgCl₂ and 0.2 mM maltose. The mixture was incubated with gentle shaking for 30 min at 22 °C, and the MBP–MalFGK₂ complex was purified by gel-filtration chromatography (Superdex 200) in buffer containing 10 mM Tris-HCl pH 8, 200 mM NaCl, 1 mM EDTA, 0.5 mM ATP, 0.2 mM maltose and 0.03% (w/v) n-undecyl-β-D-maltoside (UDM; Anatrace). For purification of selenomethionine-labelled MBP–MalFGK₂, 5 mM dithiothreitol was also included in the gel filtration buffer.

Protein samples were concentrated to 10–15 mg ml⁻¹ (Ultra-15 MWCO 100K; Millipore) in the presence of 0.06% UDM and 1.5 mM ATP. Crystals were grown by mixing protein solution with the reservoir solution containing 27% poly(ethylene glycol) 400, 500 mM NaCl, 100 mM sodium HEPES pH 7.5 and 10 mM betaine hydrochloride, at a 1:1 ratio in sitting drops by vapour diffusion at 20 °C. The crystals grew in 1–1.5 days but shattered and completely decayed within 3–4 days. Crystals were looped out of the drop and directly frozen in liquid nitrogen. X-ray diffraction data were collected at the Advanced Photon Source (APS-23ID) at 100 K. Diffraction images were processed and scaled with HKL2000 (HKL Research, Inc.)⁵⁴.

Structure determination. The protein was crystallized in space group *P*1 with one complex per unit cell (Supplementary Table 1). The structure was initially

solved through molecular replacement by PHASER (CCP4 suite)⁵⁵ using structures of unliganded MBP (PDB accession number 1OMP) and ATP-bound, closed MalK dimer (PDB accession number 1Q12). Phase improvement by density modification (CCP4 suite) produced an electron density map in which the transmembrane helices for MalF and MalG were quite evident. The model was built iteratively in O⁵⁶ and Coot⁵⁷ and refined with CNS⁵⁸. Further refinement was performed in REFMAC5 with the TLS parameters generated by the TLSMD server⁵⁹. TLS tensors were analysed and anisotropic B factors were derived with TLSANL⁶⁰. The final model, refined to 2.8 Å, contains residues 1–370 of MBP (numbered after removing the signal sequence), residues 13–242 and 245–504 of MalF, residues 7–67 and 74–296 of MalG, residues 2–371 of the MalK subunits, one maltose molecule and two ATP molecules (Supplementary Table 1). Most of the TM region has excellent density, as shown in Supplementary Fig. 1. However, the electron density map of the first TM helix of MalF shows little features for side chains, suggesting its high flexibility. This helix makes little contact with the rest of the protein, which may account for the observation that deleting this helix caused no significant effect on folding or transport function⁶¹.

Crystals of selenomethionine (SeMet)-substituted MalFGK₂(E159Q) in complex with native MBP, maltose and ATP were obtained from conditions similar to those for the native proteins, and 3.6-Å SAD data were collected at the selenium peak wavelength (Supplementary Table 1). The largest difference in the unit cell dimensions between the native and SeMet crystals is greater than 5 Å, indicating that the two crystal forms are non-isomorphous. To obtain an anomalous difference Fourier density map from SeMet data, phases were obtained from molecular replacement, with the native complex structure as a search model. Contoured at 5σ, 39 peaks were visible in the anomalous difference map; all of them co-localized with the side-chain atoms of methionine residues (Supplementary Fig. 2), confirming the correct assignment of the residues.

Figure preparation. All figures were prepared with the program PyMOL⁶². The cavity found in Fig. 5 was calculated using the program VOIDOO⁶³ with a 1.7-Å probe.

- Davidson, A. L. & Nikaido, H. Overproduction, solubilization, and reconstitution of the maltose transport system from *Escherichia coli*. *J. Biol. Chem.* **265**, 4254–4260 (1990).
- Chen, J., Lu, G., Lin, J., Davidson, A. L. & Quijcho, F. A. A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Mol. Cell* **12**, 651–661 (2003).
- Dean, D. A., Fikes, J. D., Gehring, K., Bassford, P. J. Jr & Nikaido, H. Active transport of maltose in membrane vesicles obtained from *Escherichia coli* cells producing tethered maltose-binding protein. *J. Bacteriol.* **171**, 503–510 (1989).
- Otwinski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Macro Crystallogr. A* **276**, 307–326 (1997).
- Project, C. C. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Brunker, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Painter, J. & Merritt, E. A. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Cryst.* **39**, 109–111 (2006).
- Howlin, B., Butler, S. A., Moss, D. S., Harris, G. W. & Driessen, H. P. C. TLSANL—TLS parameter-analysis program for segmented anisotropic refinement of macromolecular structures. *J. Appl. Cryst.* **26**, 622–624 (1993).
- Ehrmann, M. & Beckwith, J. Proper insertion of a complex membrane protein in the absence of its amino-terminal export signal. *J. Biol. Chem.* **266**, 16530–16533 (1991).
- DeLano, W. L. The PyMOL Molecular Graphics System (2002) on the World Wide Web (<http://www.pymol.org>).
- Kleywegt, G. J. & Jones, T. A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. D* **50**, 178–185 (1994).

LETTERS

White dwarf stars with carbon atmospheres

P. Dufour¹, J. Liebert¹, G. Fontaine² & N. Behara³

White dwarfs represent the endpoint of stellar evolution for stars with initial masses between approximately 0.07 and 8–10 M_{\odot} , where M_{\odot} is the mass of the Sun (more massive stars end their life as either black holes or neutron stars). The theory of stellar evolution predicts that the majority of white dwarfs have a core made of carbon and oxygen, which itself is surrounded by a helium layer and, for ~80 per cent of known white dwarfs, by an additional hydrogen layer^{1–3}. All white dwarfs therefore have been traditionally found to belong to one of two categories: those with a hydrogen-rich atmosphere (the DA spectral type) and those with a helium-rich atmosphere (the non-DAs). Here we report the discovery of several white dwarfs with atmospheres primarily composed of carbon, with little or no trace of hydrogen or helium. Our analysis shows that the atmospheric parameters found for these stars do not fit satisfactorily in any of the currently known theories of post-asymptotic giant branch evolution, although these objects might be the cooler counterpart of the unique and extensively studied PG 1159 star H1504+65 (refs 4–7). These stars, together with H1504+65, might accordingly form a new evolutionary sequence that follows the asymptotic giant branch.

Traces of carbon are typically observed as either neutral carbon lines or molecular C₂ Swan bands (defining the DQ spectral type) in cool helium-rich white dwarfs with effective temperatures (T_{eff}) below ~13,000 K. The presence of carbon in the atmospheres of these objects has been explained successfully by a model in which carbon is dredged-up from the underlying carbon/oxygen core by the deep helium convection zone⁸. This model predicts a maximum contamination of carbon at an effective temperature of ~10,000 K (corresponding approximately to the temperature at which the surface convection zone is maximal) before gradually decreasing with lower temperature, in agreement with atmospheric analysis determinations⁹. We note that although some of these objects show a very high level of carbon pollution, helium always remains the dominant constituent of the atmosphere (the highest carbon abundances by number (N) found are around $\log N(\text{C})/N(\text{He}) \approx -3$).

Atmospheric abundance determination for DQ white dwarfs has been most successful for stars on the cool side of the maximum contamination and very few analyses have been done for hotter objects located on the ascending side of the curve^{10–12}. Recently¹³, several new hot DQ white dwarfs have been discovered in the Sloan Digital Sky Survey (SDSS), providing a unique opportunity to test the dredge-up theory on the ‘hot side’. However, appropriate atmospheric models were not available at that time. Because the atmospheric compositions of these white dwarfs were expected to be helium-dominated, effective temperature estimates (T_{He}) were obtained by comparing the SDSS colours with those of pure helium composition models. Although it was expected that such an approximation would overestimate the effective temperature by several thousand kelvin, some of these stars were so hot ($T_{\text{He}} > 25,000$ K) that an overlap with most of the helium-rich star (the DB spectral type) temperature range seemed unavoidable. It was hypothesized

that these stars had thinner outer helium envelopes so that dredge-up occurred earlier in the cooling sequence. These highly carbon-polluted white dwarfs are expected to be massive, so it was argued that they might represent the missing high-mass tail of the DB mass distribution¹⁴.

Thus, it is with this scientific rationale in mind that we proceeded with the calculation of the appropriate atmospheric models for these objects. Because the continuum opacity of heavy elements might not be negligible for these objects, these new models have been updated with the latest C and O photoionization cross-sections from the Opacity Project¹⁵. Although the analysis of the coolest objects ($T_{\text{eff}} < 15,000$ K) is straightforward (results will be presented elsewhere; manuscript in preparation), we quickly realized that no combination of carbon and helium could successfully reproduce the observed features (mostly C II lines) in the optical spectrum of the hottest ones by assuming a helium-dominated atmosphere. Indeed, such models predict the presence of a strong He I $\lambda = 4,471$ line that is not observed spectroscopically in our sample of hot DQ stars. We thus concluded that a good fit to both the spectra and the energy distribution was possible only by considering atmospheres made primarily of carbon, with little or no trace of hydrogen and helium.

Figure 1 shows our fits to the optical spectrum and photometric energy distribution of one such star using a grid of pure carbon atmosphere models, as well as models containing traces of hydrogen and helium. The fitting method is similar to that used for cooler DQ and need not be repeated here⁹. This particular case shows a small trace of hydrogen which we believe is more likely to be the result of accretion from the interstellar medium than of primordial origin. We found eight more similar objects in the SDSS white dwarf catalogue¹⁶, all of which were found to have a carbon-dominated atmosphere and a temperature between 18,000 and 23,000 K, although for some faint stars detailed analysis is more uncertain because of the poor signal-to-noise ratio of the observations. Only crude upper limits for hydrogen and helium (of the order of $\log N(\text{C})/N(\text{H}) > 1.5$ and $\log N(\text{C})/N(\text{He}) > 1.5$) can be obtained from the other spectra. Our detailed model atmosphere fits provide fair estimates of the effective temperature and chemical composition, but higher-quality data (which we hope to obtain in December 2007, weather permitting) are essential for a precise measure of the surface gravity (and thus mass) of these objects.

These new stars are too hot to be explained by the standard convective dredge-up scenario and so we sought another explanation. Natural progenitor candidates to consider are the hot PG 1159 stars. The latter are probably the result of a late He-shell flash¹⁷ (the so-called ‘born-again’ scenario) at the end of the post-asymptotic-giant-branch phase, which almost completely eliminates the remaining hydrogen and mixes material from the interior with the helium envelope. As a result, these objects re-enter the white dwarf cooling track, but this time they have a surface composition that is a mixture of helium, carbon, oxygen and little or no hydrogen (T_{eff} between ~75,000 and 200,000 K; typical abundances¹⁸, in mass fraction, of

¹Department of Astronomy and Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, Arizona 85721, USA. ²Département de Physique, Université de Montréal, CP 6128, Succursale Centre-Ville, Montréal, QC H3C 3J7, Canada. ³CIFIST, GEPI, Observatoire de Paris-Meudon, 92195, France.

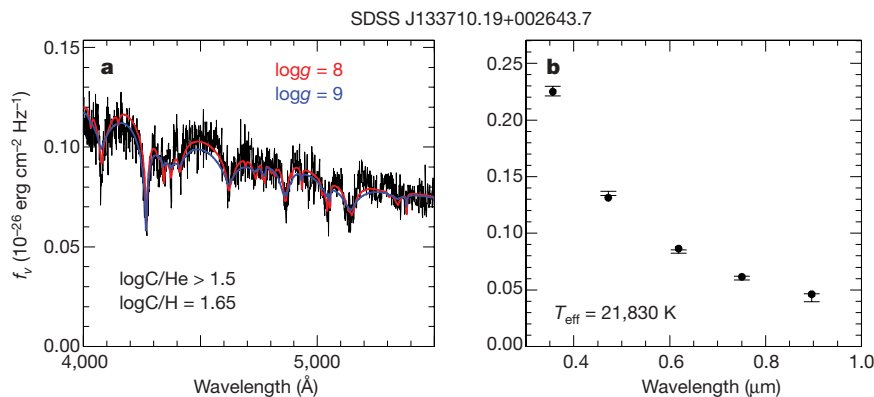


Figure 1 | Fit to the optical spectra and energy distribution for a carbon-rich white dwarf. **a**, The thick red line represents our $\log g = 8$ best fit (the parameters are indicated in the panel), whereas the blue line represents a solution for $\log g = 9$. Hydrogen abundance is determined by fitting H β ($\lambda = 4,861$ Å), whereas helium abundance is constrained by the absence of

He, C, O and Ne are 33%, 50%, 15% and 2%). However, as these stars cool down, gravitational diffusion rapidly separates the heavier elements from the helium that tends to float to the surface, and, by the time they have reached 25,000 K, helium completely dominates the surface composition.

Models exploring the evolution of 9–11 M_{\odot} stars have produced white dwarfs with O–Ne–Mg cores and CO envelopes and possibly little He and H; such models may explain our peculiar stars^{19–21}. However, it is unclear what quantity of He and H should remain for such stars. Furthermore, the resulting white dwarfs have high surface gravities ($\log g \approx 9$, in centimetre-gram-second units) and would thus show extremely broad carbon lines that are incompatible with the observed widths of the lines (Fig. 1 shows that $\log g \approx 8$ is more likely, although analysis of a better spectrum is needed for more precision). The most likely explanation for the existence of hot DQ white dwarfs with carbon-dominated atmospheres is that these stars are the progenies of objects such as H1504+65. The latter is a unique object among the known PG 1159 stars so far. It is the hottest specimen of its class at $T_{\text{eff}} \approx 200,000$ K, and its atmospheric composition is very unusual, with a mass fraction of $\sim 50\%$ C and $\sim 50\%$ O plus small traces of heavier elements, but no detectable helium or hydrogen. It is currently believed that H1504+65 is essentially a bare stellar nucleus produced by a particularly violent post-asymptotic-giant-branch very-late thermal pulse that has destroyed in large part the remaining stellar envelope containing helium and hydrogen. The best available simulations of such an event, those of ref. 17, suggest indeed that, following the last He flash, the full envelope becomes convective and extends to deep enough layers for H and He to be completely consumed, except that very small traces of He may survive.

The uniformity of the chemical composition of the envelope of H1504+65 is believed to be maintained for a while by a residual wind but, with time, diffusion becomes dominant and He, C and O must separate under the influence of gravitational settling. The idea is that the wind slowly dies out with cooling (decreasing luminosity) and that the braking effect it provides on element separation becomes less and less efficient. Hence, the small residual trace of helium believed to exist in the envelope of H1504+65 eventually diffuses upward to form a thin layer above a C-enriched and O-depleted mantle. The total mass of an atmosphere is tiny ($\sim 10^{-14}$ to $10^{-15} M_{\star}$, where M_{\star} is the mass of the star), so there is ultimately enough accumulated He to form a full atmosphere and the descendant of H1504+65 would now be ‘disguised’ as a He-atmosphere white dwarf after the PG1159 evolutionary phase. With further cooling, a convection zone develops in the C-enriched mantle owing to the recombination of that element, while the overlying He layer remains radiative. At some T_{eff} , the exact value of which is not yet known owing to a current lack of

the He I $\lambda = 4,471$ Å line. **b**, Photometric measurements in the u, g, r, i and z bands are represented by error bars, whereas the average model fluxes from the same best-fit $\log g = 8$ model are shown by filled circles. Both **a** and **b** have the same y axis.

proper models, it is conjectured that the subphotospheric C convection zone becomes active enough to be able to dilute from below the overlying He layer. At that point the star would undergo a dramatic spectral change, transforming itself from a He-dominated atmosphere white dwarf to a star with a C-dominated atmosphere, because the mass in the C convection zone is orders of magnitude larger than the mass of the He layer.

Hence, the former PG 1159 star H1504+65, showing initially a mixed C and O atmosphere, would now show a C-dominated atmosphere after an intermediate phase in which it would have been observed as a He-atmosphere white dwarf. We note, however, that helium must ultimately reappear at the surface in enough quantity to form a helium-rich atmosphere because no carbon-rich object has ever been discovered at lower temperatures ($T_{\text{eff}} < 15,000$ K). The remaining helium probably floats again at the surface when the star cools down, perhaps turning the carbon-rich objects into DQ stars belonging to the second sequence of high carbon abundances^{11,22}. We are at present making evolutionary calculations to test the above scenario.

To conclude, we roughly estimate that the space density of carbon-rich white dwarfs, assuming a 50% completeness for the targeting in SDSS¹⁶, is between $\sim 2.1 \times 10^{-6} \text{ pc}^{-3}$ and $\sim 7.0 \times 10^{-8} \text{ pc}^{-3}$ depending on the preferred value of $\log g$ (and thus radius) and T_{eff} for the stars. For comparison, the local space density of white dwarfs²³ is estimated to be $\sim 5.0 \times 10^{-3} \text{ pc}^{-3}$. White dwarfs with carbon-dominated atmospheres are thus intrinsically rare and a relatively large volume of space (as was done by the Sloan Digital Sky Survey) had to be surveyed before many of them could be found.

Received 31 July; accepted 12 September 2007.

1. Iben, I. Jr. On the frequency of planetary nebula nuclei powered by helium burning and on the frequency of white dwarfs with hydrogen-deficient atmospheres. *Astrophys. J.* **277**, 333–354 (1984).
2. Koester, D. & Schoenberner, D. Evolution of white dwarfs. *Astron. Astrophys.* **154**, 125–134 (1986).
3. D’Antona, F. & Mazzitelli, I. Evolutionary times of white dwarfs—Long or short? *IAU Colloq.* **95**, 635–637 (1987).
4. Nousek, J. A. et al. H 1504 + 65—an extraordinarily hot compact star devoid of hydrogen and helium. *Astrophys. J.* **309**, 230–240 (1986).
5. Werner, K. NLTE analysis of the unique pre-white dwarf H 1504 + 65. *Astron. Astrophys.* **251**, 147–160 (1991).
6. Werner, K. & Wolff, B. The EUV spectrum of the unique bare stellar core H1504+65. *Astron. Astrophys.* **347**, L9–L13 (1999).
7. Werner, K., Rauch, T., Barstow, M. A. & Kruk, J. W. Chandra and FUSE spectroscopy of the hot bare stellar core H 1504+65. *Astron. Astrophys.* **421**, 1169–1183 (2004).
8. Pelletier, C., Fontaine, G., Wesemael, F., Michaud, G. & Wegner, G. Carbon pollution in helium-rich white dwarf atmospheres. Time-dependent calculations of the dredge-up process. *Astrophys. J.* **307**, 242–252 (1986).

9. Dufour, P., Bergeron, P. & Fontaine, G. Detailed spectroscopic and photometric analysis of DQ white dwarfs. *Astrophys. J.* **627**, 404–417 (2005).
10. Wegner, G. & Koester, D. Atmospheric analysis of the carbon white dwarf G227–5. *Astrophys. J.* **288**, 746–750 (1985).
11. Thejll, P., Shipman, H. L., MacDonald, J. & Macfarland, W. M. An atmospheric analysis of the carbon-rich white dwarf G35 – 26. *Astrophys. J.* **361**, 197–206 (1990).
12. Desharnais, S., Wesemael, F., Chayer, P. & Kruk, J. W. FUSE observation of cool DB white dwarfs. *ASP Conf. Ser.* **372**, 265–268 (2007).
13. Liebert, J. et al. SDSS white dwarfs with spectra showing atomic oxygen and/or carbon lines. *Astron. J.* **126**, 2521–2528 (2003).
14. Beauchamp, A. *Détermination des Paramètres Atmosphériques des Étoiles Naines Blanches de Type DB*. PhD thesis, Montréal (1995).
15. Behara, N. & Jeffery, C. S. LTE model atmosphere with new opacities. 1. Methods and general properties. *Astron. Astrophys.* **451**, 643–650 (2006).
16. Eisenstein, D. J. et al. A catalog of spectroscopically confirmed white dwarfs from the Sloan Digital Sky Survey data release 4. *Astrophys. J.* **167** (Suppl.), 40–58 (2006).
17. Herwig, F., Bloeker, T., Langer, N. & Driebe, T. On the formation of hydrogen-deficient post-AGB stars. *Astron. Astrophys.* **349**, L5–L8 (1999).
18. Werner, K. & Herwig, F. The elemental abundances in bare planetary nebula central stars and the shell burning in AGB stars. *Publ. Astron. Soc. Pacif.* **118**, 183–204 (2006).
19. Garcia-Berro, E. & Iben, I. On the formation and evolution of super-asymptotic giant branch stars with cores processed by carbon burning. 1: SPICA to Antares. *Astrophys. J.* **434**, 306–318 (1994).
20. Garcia-Berro, E., Ritossa, C. & Iben, I. On the evolution of stars that form electron-degenerate cores processed by carbon burning. III. The inward propagation of a carbon-burning flame and other properties of a $9M_{\odot}$ model star. *Astrophys. J.* **485**, 765–784 (1997).
21. Ritossa, C., Garcia-Berro, E. & Iben, I. On the evolution of stars that form electron-degenerate cores processed by carbon burning. V. Shell convection sustained by helium burning, transient neon burning, dredge-out, URCA cooling, and other properties of an $11M_{\odot}$ population I model star. *Astrophys. J.* **515**, 381–397 (1999).
22. Koester, D. & Knist, S. New DQ white dwarfs in the Sloan Digital Sky Survey DR4: confirmation of two sequences. *Astron. Astrophys.* **454**, 951–956 (2006).
23. Holberg, J. B., Oswalt, T. D. & Sion, E. M. A determination of the local density of white dwarf stars. *Astrophys. J.* **571**, 512–518 (2002).

Acknowledgements This work has been partially supported by the NSF for work on SDSS white dwarfs. This work was also supported in part by the NSERC (Canada).

Author Contributions P.D. developed the models, performed the analysis and wrote the paper. J.L. was involved in the discovery of the SDSS spectra. G.F. elaborated the evolutionary scenario discussed in the text. N.B. provided opacity data necessary to the computation of the models. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.D. (dufourpa@as.arizona.edu).

Coupled ^{142}Nd – ^{143}Nd evidence for a protracted magma ocean in Mars

V. Debaille¹, A. D. Brandon², Q. Z. Yin³ & B. Jacobsen³

Resolving early silicate differentiation timescales is crucial for understanding the chemical evolution and thermal histories of terrestrial planets¹. Planetary-scale magma oceans are thought to have formed during early stages of differentiation, but the longevity of such magma oceans is poorly constrained. In Mars, the absence of vigorous convection and plate tectonics has limited the scale of compositional mixing within its interior², thus preserving the early stages of planetary differentiation. The SNC (Shergotty–Nakhla–Chassigny) meteorites from Mars retain ‘memory’ of these events^{3–5}. Here we apply the short-lived ^{146}Sm – ^{142}Nd and the long-lived ^{147}Sm – ^{143}Nd chronometers to a suite of shergottites to unravel the history of early silicate differentiation in Mars. Our data are best explained by progressive crystallization of a magma ocean with a duration of ~100 million years after core formation. This prolonged solidification requires the existence of a primitive thick atmosphere on Mars that reduces the cooling rate of the interior⁶.

Metallic core formation in Mars is constrained to 7–15 Myr after solar nebula condensation from ^{182}Hf – ^{182}W chronometry on SNC meteorites^{5,7}. Core formation is thought to occur in conjunction with planetary-scale magma oceans that result from melting in the final stages of accretion⁸. Accretion, radioactive decay of short-lived isotopes such as ^{26}Al , and gravitational energy from core formation provided enough heat to melt a large portion of the martian mantle¹. The rate of magma ocean solidification remains poorly constrained. Time intervals of magma ocean crystallization for Earth- and Mars-sized bodies from several thousand to a few hundred millions of years have been proposed^{1,6,9}.

Sm and Nd fractionate from each other during silicate differentiation because Nd is more incompatible than Sm during magmatic processes, resulting in greater Sm/Nd in residues than in partial melts. The ^{146}Sm – ^{142}Nd chronometer with a half-life of 103 million years (Myr), coupled with the ^{147}Sm – ^{143}Nd system with a half-life of 106 billion years (Gyr), can track early silicate differentiation events during extant ^{146}Sm of ~500 Myr. Anomalies in $^{142}\text{Nd}/^{144}\text{Nd}$ have previously been reported for SNCs^{3,5,10–13}. These data are consistent with one or more silicate differentiation events at $4,525^{+19}_{-21}$ Myr ago, close to the time of core formation⁵. Variation in $^{142}\text{Nd}/^{144}\text{Nd}$ and the calculated $^{143}\text{Nd}/^{144}\text{Nd}$ for the mantle source compositions of SNCs presumably resulted from rapid crystallization of a martian magma ocean (MMO) that produced compositionally distinct materials within the mantle⁴, and imply timescales of solidification of 40 Myr or less. However, the range of uncertainties in $^{142}\text{Nd}/^{144}\text{Nd}$ of ± 12 – 31 p.p.m. (2σ) (1 p.p.m. corresponds to 10^{-6}) permits differentiation times as early as 4,567 Myr to younger than 4,000 Myr ago for individual rocks. Hence, the relationship between each rock reservoir and a single early magma ocean, the duration of this magma ocean, or whether there were multiple planetary-scale silicate

differentiation events during the lifetime of ^{146}Sm , all remain unclear. In addition, earlier assessment was made assuming that the terrestrial standard value for $^{142}\text{Nd}/^{144}\text{Nd}$ was chondritic and represented the starting composition of the martian mantle. Precise $^{142}\text{Nd}/^{144}\text{Nd}$ measurements on present-day mantle-derived Earth samples show a clearly resolved difference with chondrites resulting from long-term differences in their Sm/Nd (refs 14–16). These differences affect the calculated ages derived from multi-stage mantle evolution models.

Here we report Sm–Nd isotope systematics for eight martian shergottites that span the range of known compositions. Neodymium from the samples was purified using cation column chromatography and Ln-Spec resin (di(2-ethylhexyl) orthophosphoric acid (HDEHP) saturated on teflon beads). Cerium was removed from the Nd cuts by solvent extraction¹⁷. The high-precision Nd isotope measurements were made on a ThermoFinnigan Triton thermal ionization mass spectrometer as positive metal ions in multidynamic mode. Details of analytical techniques are presented in the Methods and in the Supplementary Information. The $^{142}\text{Nd}/^{144}\text{Nd}$ was measured to precisions of ± 2 – 6 p.p.m. (2σ) on these samples, allowing for a more precise evaluation of the timescales and number of early silicate differentiation events. The Nd isotope data are presented in ϵ notation (Fig. 1, Table 1), defined as deviation in parts per 10^4 from the

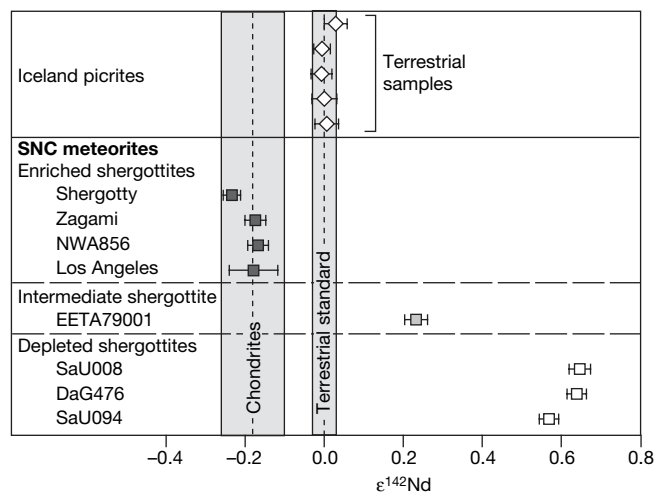


Figure 1 | $\epsilon^{142}\text{Nd}$ measured for martian meteorites in this study. Terrestrial samples (Iceland picrites) measured during the same analytical campaign are added for comparison. $\epsilon^{142}\text{Nd}$ is defined as:

$$\epsilon^{142}\text{Nd} = \left(\frac{(^{142}\text{Nd}/^{144}\text{Nd})_{\text{sample}}}{(^{142}\text{Nd}/^{144}\text{Nd})_{\text{terrestrial standard}}} - 1 \right) \times 10,000.$$

Shaded areas represent the chondrite value (-0.18 ± 0.08^{15}), and the terrestrial standard (0.0 ± 0.03). Error bars represent $\pm 2\sigma$ for each measurement.

¹Lunar and Planetary Institute, 3600 Bay Area Boulevard, Houston, Texas 77058, USA. ²NASA Johnson Space Center, Mail Code KR, Houston, Texas 77058, USA. ³Department of Geology, University of California Davis, One Shield Avenue, Davis, California 95616, USA.

terrestrial standard for $\epsilon^{142}\text{Nd}$ and the chondrite uniform reservoir (CHUR) for $\epsilon^{143}\text{Nd}$ (ref. 14).

The average chondrite value for $\epsilon^{142}\text{Nd}$ is -0.18 ± 0.08 (ref. 15), while that for the terrestrial standard is 0 ± 0.03 (refs 14–16) (Fig. 1). The depleted shergottites SaU008, SaU094, and DaG476 have $\epsilon^{142}\text{Nd}$ values from $+0.57 \pm 0.04$ to $+0.65 \pm 0.05$ (2σ), and initial $\epsilon^{143}\text{Nd}$ values from $+36.2$ to $+39.1$, indicating derivation from an incompatible-trace-element-depleted source formed during the first 500 Myr evolution of Mars. The intermediate (EETA79001) and enriched shergottites (Shergotty, Zagami, Los Angeles, NWA856) show a progressive decrease in $^{142}\text{Nd}/^{144}\text{Nd}$ and $^{143}\text{Nd}/^{144}\text{Nd}$, from incompatible-trace-element-depleted to enriched with $\epsilon^{142}\text{Nd}$ from $+0.23 \pm 0.03$ to -0.23 ± 0.02 , and initial $\epsilon^{143}\text{Nd}$ from $+16.9$ to -7.0 , respectively.

The evolution of source reservoirs for shergottites is modelled with a bulk Mars that is initially chondritic for Sm/Nd with present-day $\epsilon^{142}\text{Nd} = -0.18$ and $\epsilon^{143}\text{Nd} = 0$. The $\epsilon^{142}\text{Nd}$ and $\epsilon^{143}\text{Nd}$ for each shergottite is calculated using a two-stage model (Fig. 2), in which $\epsilon^{143}\text{Nd}$ is calculated from $^{147}\text{Sm}/^{144}\text{Nd}$ ratios for time-integrated sources at t_1 and projected in time to 150 Myr ago:

$$\left(\frac{^{142}\text{Nd}}{^{144}\text{Nd}}\right)_{t_2} = \left(\frac{^{142}\text{Nd}}{^{144}\text{Nd}}\right)_{\text{present}}^{\text{CHUR}} - \left(\frac{^{146}\text{Sm}}{^{144}\text{Sm}}\right)_{T_0}^{\text{CHUR}} \times \left(\frac{^{144}\text{Sm}}{^{147}\text{Sm}}\right)_{\text{present}}^{\text{CHUR}} \times \left[\left(\frac{^{147}\text{Sm}}{^{144}\text{Nd}}\right)_{t_1}^{\text{CHUR}} \times e^{-\lambda_{146}(T_0 - t_1)} + \left(\frac{^{147}\text{Sm}}{^{144}\text{Nd}}\right)_{t_1}^{\text{source}} \times \left(e^{-\lambda_{146}(T_0 - t_2)} - e^{-\lambda_{146}(T_0 - t_1)} \right) \right] \quad (1)$$

$$\left(\frac{^{143}\text{Nd}}{^{144}\text{Nd}}\right)_{t_2}^{\text{source}} = \left(\frac{^{143}\text{Nd}}{^{144}\text{Nd}}\right)_{\text{present}}^{\text{CHUR}} + \left(\frac{^{147}\text{Sm}}{^{144}\text{Nd}}\right)_{\text{present}}^{\text{CHUR}} \times (1 - e^{\lambda_{147}t_1}) + \left(\frac{^{147}\text{Sm}}{^{144}\text{Nd}}\right)_{t_1}^{\text{source}} \times (e^{\lambda_{147}t_1} - e^{\lambda_{147}t_2}) \quad (2)$$

where $T_0 = 4.567$ Gyr ago, the onset of solar nebula condensation, t_1 is the time of Sm/Nd fractionation, and t_2 is the crystallization age of each meteorite (Table 1). $\lambda_{146} = 6.74 \times 10^{-9} \text{ yr}^{-1}$ and $\lambda_{147} = 6.54 \times 10^{-12} \text{ yr}^{-1}$ are the decay constants for ^{146}Sm and ^{147}Sm respectively. In this two-stage model, samples plotting inside the field defined by isochrons for t_1 (Fig. 2) give information on the timing of shergottite source formation. Shergottites plot inside the isochron field, and define a line with an R^2 coefficient of 0.997. It has been proposed that this linear array represents both a mixing line and an isochron, implying near-simultaneous formation of enriched and depleted shergottite end-members^{5,10,12,18,19}. Our new high-precision data show that this line misses the origin point (Fig. 2) and is not an isochron. Instead, the coupled ^{142}Nd – ^{143}Nd systematics for the depleted and enriched shergottites constrain earlier and later Sm/Nd fractionation events, respectively. Thus, the formation age of the enriched end-member is not simultaneous with the depleted shergottite source. This observation still holds if Mars, instead of being initially chondritic, had an initial composition similar to that of the Earth's depleted mantle, characterized by present-day $\epsilon^{142}\text{Nd} = 0$ and $\epsilon^{143}\text{Nd} = +10.69$ (ref. 16). The shergottite mixing line still misses the origin point and is independent of these starting parameters.

A straightforward interpretation of the two-stage model for the depleted shergottite source is that it marks the time for early-formed cumulates in an MMO⁴. Solving equations (1) and (2) for depleted shergottites leads to a source formation age of $4,535_{-7}^{+7}$ Myr ago, which is ~ 32 Myr after Solar System condensation (Fig. 2).

The origin of the enriched shergottites is at present debated, and two scenarios may explain the observed coupled ^{142}Nd – ^{143}Nd systematics. One (scenario A) is that the depleted shergottite source underwent variable and progressive enrichment at some time after

development of this source but before production of the first shergottite magma at 474 Myr ago (Table 1). This could result from migration of metasomatic fluids or silicate melts in the martian mantle^{12,20}. A second (scenario B) is that the regression line corresponds to mixing between a depleted source and a distinct enriched source that formed separately. The distinct enriched source could be late-stage quenched residual melt from the MMO formed within the martian mantle^{4,12} or the martian crust^{10,19,21,22}. This enriched source remains separated from the depleted reservoir until the time of melting that results in shergottite magmatism.

The shergottites plot as a mixing line to the right of a 1:1 line for time-integrated $^{147}\text{Sm}/^{144}\text{Nd}$ for their sources versus those measured in the lavas (Fig. 3). This means that the $^{147}\text{Sm}/^{144}\text{Nd}$ source ratio calculated from the initial $^{143}\text{Nd}/^{144}\text{Nd}$ at the time of crystallization for each rock is less than that measured in the lavas, which is the opposite of what occurs during partial melting. This can be explained by a first episode of partial melt extraction (first stage) in the shergottite source just before a second episode of partial melting (second stage) that produced shergottite magma (ref. 10; Supplementary Fig. 2). The first stage results in increasing Sm/Nd in the residue because Nd is more incompatible than Sm. The second-stage shergottite magma produced from the first-stage residue has Sm/Nd > initial Sm/Nd before first-stage partial melt extraction and plots to the right of the 1:1 line (Fig. 3). Because the $^{143}\text{Nd}/^{144}\text{Nd}$ in the source does not have time to evolve between these two events¹⁰, the calculated source Sm/Nd is that before the first-stage partial melting, and is thus lower than the Sm/Nd measured. Shergottites have a range of crystallization ages from 165 to 474 Myr ago (Table 1). If their parental magmas issued from a single depleted source that underwent progressive degrees of metasomatism as in scenario A above, then the cycle of first-stage melting that predates each second stage of shergottite magma production was repeated multiple times. Given these systematics, scenario A requires that variable degrees of partial melting in multiple melting stages and variable degrees of metasomatism result in measured and calculated source Sm/Nd for each shergottite in a

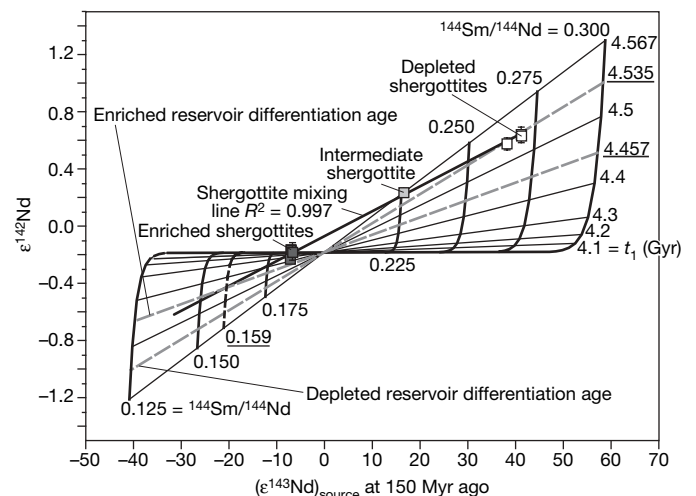


Figure 2 | A two-stage coupled ^{142}Nd – ^{143}Nd evolution model for a chondritic martian magma ocean projected to 150 Myr ago. Symbols and error bars (comprised in symbol size) as in Fig. 1. The black bold line represents the mixing trend between all shergottites and misses the origin point. Black lines are loci of equal differentiation ages. Dashed grey lines are the differentiation ages for depleted and enriched shergottite sources, respectively. Black bold curves are loci of equal $^{147}\text{Sm}/^{144}\text{Nd}$ ratios in the source, which are identical at t_1 and the present-day. The dashed bold curve represents the maximum $^{147}\text{Sm}/^{144}\text{Nd}$ value of 0.159 (underlined) for the enriched end-member. Parameters used in calculations are: $(^{147}\text{Sm}/^{144}\text{Nd})_{\text{CHUR}}^{\text{CHUR}} = 0.1966$; $(^{143}\text{Nd}/^{144}\text{Nd})_{\text{CHUR}}^{\text{CHUR}} = 0.512638$; $(^{146}\text{Sm}/^{144}\text{Sm})_{T_0} = 0.008$ (see discussion in ref. 27); $\epsilon^{142}\text{Nd}_{\text{CHUR}} = -0.18$; and $(^{147}\text{Sm}/^{144}\text{Sm})_{\text{CHUR}} = 4.88899$.

Table 1 | Sm–Nd isotope values for shergottites

	Crystallization age t_2 (Myr)	Measured $^{147}\text{Sm}/^{144}\text{Nd}$	$\epsilon^{142}\text{Nd}$ (2σ)	$\epsilon^{143}\text{Nd}$ at t_2
Enriched shergottites				
Shergotty	165 (ref. 28)	0.2250	−0.233	±0.022
Zagami	177 (ref. 28)	0.2247	−0.174	±0.026
NWA856	186 (ref. 29)	0.2218	−0.167	±0.026
Los Angeles	170 (ref. 28)	0.2224	−0.179	±0.061
Intermediate shergottite				
EETA79001	173 (ref. 28)	0.3971	0.233	±0.029
Depleted shergottites				
SaU008	446 (ref. 30)	0.4863	0.647	±0.054
DaG476	474 (ref. 28)	0.4924	0.639	±0.031
SaU094	446 (ref. 30)	0.4629	0.569	±0.042

$^{147}\text{Sm}/^{144}\text{Nd}$ are the directly measured values in the samples. $\epsilon^{143}\text{Nd}$ are calculated at t_2 (age of crystallization) relative to CHUR. See Supplementary information for more details.

manner that forms a well-defined line. This complex scenario is fortuitous and highly unlikely.

A simpler explanation is scenario B, in which all shergottites are derived from a unique depleted source having first undergone one partial melting event before shergottite parental magma production. The enriched and intermediate shergottites are then explained by mixing different proportions of melts derived from the depleted source and a distinct enriched source. In this case, the enriched source has Sm/Nd smaller than the most enriched measured lava, and falls to the left of the 1:1 line. This corresponds to $^{147}\text{Sm}/^{144}\text{Nd} \leq 0.159$ (Fig. 3). For this $^{147}\text{Sm}/^{144}\text{Nd}$ value, the enriched source formed at 4,457 Myr ago for a chondritic Mars (Fig. 2) or ~ 100 Myr after core formation. These Sm–Nd systematics could reflect either the final stages of MMO solidification, or the crust that contaminates the depleted shergottite magmas.

The well-defined mixing lines projecting away from the depleted shergottites in Figs 2 and 3 are inconsistent with the scatter expected where individual magma batches are contaminated by crust. This is supported by the chondritic initial $^{187}\text{Os}/^{188}\text{Os}$ of all shergottites²³. Ancient basaltic or felsic crust on Mars will have superchondritic Re/Os and evolve to very high $^{187}\text{Os}/^{188}\text{Os}$ in hundreds of Myr, such that very small amounts of crust would drive contaminated magmas to much greater $^{187}\text{Os}/^{188}\text{Os}$ ratios than observed²³. Also, NWA1068 is an enriched shergottite with a more primitive major-element composition (that is, molar (Mg/Mg+Fe) = 0.59) than those studied here (molar (Mg/Mg+Fe) = 0.25–0.52), but they all have nearly identical $\epsilon^{143}\text{Nd}$ values of −6.5 to −7.0 (Table 1; ref. 24). This is inconsistent with models for progressive crustal contamination during fractional crystallization²⁰.

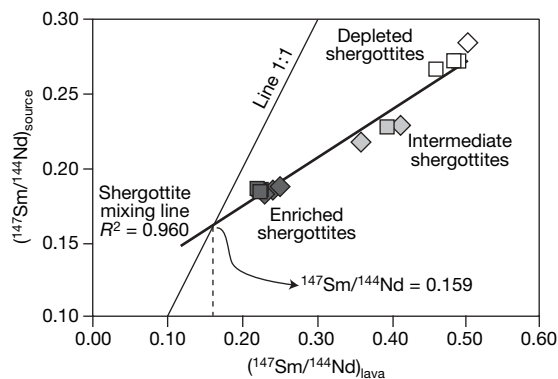


Figure 3 | The $^{147}\text{Sm}/^{144}\text{Nd}$ time-integrated ratios for mantle sources versus the measured ratios in lavas. The $^{147}\text{Sm}/^{144}\text{Nd}$ source ratio calculated using equation (2) and considering that the source differentiated at T_0 and evolved with constant parent/daughter ratios until the sample being measured was extracted from the mantle at t_2 . All of the shergottites plot to the right of the 1:1 line. This has been explained by two partial melting stages close in time in the shergottite source. Data from this study: square symbols as in Fig. 1. Data from the literature^{10,20,31,32}; diamond symbols, with the same greyscale as the data from this study.

In summary, the linear arrays observed for Sm–Nd isotope systematics of the shergottites are consistent with mixing between two distinct source reservoirs formed at ~ 4.535 Gyr and ~ 4.457 Gyr ago. The origin of the incompatible-trace-element-enriched reservoir is not likely to be found in ancient crust or metasomatism of the mantle. Instead, the most probable origin for this reservoir is late-stage quenched residual melt from the MMO. For a $^{147}\text{Sm}/^{144}\text{Nd}$ ratio between 0.150 and 0.159, comparable to the value of 0.164 inferred for the lunar KREEP (that is, strongly enriched in potassium (K), rare-earth elements (REE) and phosphorus (P)) component²⁵ and to the value of 0.152–0.154 for the trapped melt at the end of MMO crystallization^{4,10}, the formation time interval of the depleted and enriched reservoirs record progressive MMO crystallization to ~ 100 Myr after core formation. This is in agreement with the survival of shallow magma oceans for 100–200 Myr without any heating processes in the presence of a transient atmosphere⁶. It implies that a primitive atmosphere was present on Mars during this time interval, first acting as an insulating blanket and slowing down the solidification process⁶, but later removed within the first billion years²⁶.

METHODS SUMMARY

Coarse fractions of three desert meteorites (Los Angeles, SaU008 and DaG476) were leached in 50% acetic acid (see ‘Terrestrial contamination’ in the Supplementary Information). Powders were dissolved in Teflon beakers using ultrapure and concentrated HF–HNO₃–HClO₄ (10:2:1) at 160 °C for 1 week. After evaporation, concentrated HCl was added and residues were redissolved at 160 °C for 1 week. After re-dissolution in HCl, a 5% aliquot of each solution was spiked with a mixed ^{150}Nd – ^{149}Sm spike. Rare-earth elements (REE) were purified by cation exchange chromatography using AG50-X8 resin, followed by Ce, Sm and Nd separation with Eichrom Ln-Spec resin. The Nd cut was then purified of Ce using solvent extraction¹⁷ (see ‘Analytical techniques’ in the Supplementary Information). Total procedural blanks are ≤ 30 pg and ≤ 60 pg for Nd and Sm, respectively.

Spiked Sm and Nd aliquots were analysed on a multi-collector inductively coupled plasma mass spectrometer (MC-ICP-MS, Nu Plasma) at the University of California in Davis. The high-precision $^{142}\text{Nd}/^{144}\text{Nd}$ measurements of unspiked Nd aliquots were performed on a nine-faraday-cup Triton thermal ionization mass spectrometer at the Johnson Space Center (Supplementary Tables 1 and 2), as positive metal ions in multidynamic mode with rotating amplifiers between blocks²⁷. The data are normalized to $^{146}\text{Nd}/^{144}\text{Nd} = 0.7219$ to correct for instrument mass fractionation. The interference of ^{142}Ce on ^{142}Nd was monitored with ^{140}Ce . Corrections for this interference to $^{142}\text{Nd}/^{144}\text{Nd}$ ranged from 0.2 to 9.5 p.p.m., and for all but one standard run and four Iceland picrite runs, were ≤ 4 p.p.m. and within 2σ internal and established 2σ external precisions on $^{142}\text{Nd}/^{144}\text{Nd}$ (Supplementary Table 2). Sm interferences on Nd were monitored with ^{147}Sm , which was never detectable on the faraday cups above background noise. Scans on the electron multiplier showed that ^{147}Sm was never above 1,000 counts s^{-1} , and therefore negligible. Internal precisions of ± 3 p.p.m. were achieved for 600 ng standard aliquots for measuring times of 5 to 7 h. The average for these standards was $^{142}\text{Nd}/^{144}\text{Nd} = 1.1418402 \pm 0.0000034$ ($n = 14$, 2σ) and is used as the reference value of $\epsilon^{142}\text{Nd} = 0$ for calculating $\epsilon^{142}\text{Nd}$. The 2σ external precision for these 14 standard runs for $\epsilon^{142}\text{Nd}$ is ± 0.03 (± 3 p.p.m.) (see ‘Mass spectrometry’ in Supplementary Information).

Received 16 May; accepted 24 September 2007.

- Elkins-Tanton, L. T., Hess, P. C. & Parmentier, E. M. Possible formation of ancient crust on Mars through magma ocean processes. *J. Geophys. Res.* **110**, E12501, doi:10.1029/2005JE002480 (2005).
- Kiefer, W. S. Melting in the martian mantle: Shergottite formation and implications for present-day mantle convection on Mars. *Meteorit. Planet. Sci.* **38**, 1815–1832 (2003).
- Harper, C. L., Nyquist, L. E., Bansal, B., Wiesmann, H. & Shih, C.-Y. Rapid accretion and early differentiation of Mars indicated by $^{142}\text{Nd}/^{144}\text{Nd}$ in SNC meteorites. *Science* **267**, 213–217 (1995).
- Borg, L. E. & Draper, D. S. A petrogenetic model for the origin and compositional variation of the martian basaltic meteorites. *Meteorit. Planet. Sci.* **38**, 1713–1731 (2003).
- Foley, N. C. *et al.* The early differentiation history of Mars from ^{182}W - ^{142}Nd isotope systematics in the SNC meteorites. *Geochim. Cosmochim. Acta* **69**, 4557–4571 (2005).
- Abe, Y. Thermal and chemical evolution of the terrestrial magma ocean. *Phys. Earth Planet. Inter.* **100**, 27–39 (1997).
- Kleine, T., Mezger, K., Münker, C., Palme, H. & Bischoff, A. ^{182}Hf - ^{182}W isotope systematics of chondrites, eucrites, and martian meteorites: chronology of core formation and early mantle differentiation in Vesta and Mars. *Geochim. Cosmochim. Acta* **68**, 2935–2946 (2004).
- Lee, D.-C. & Halliday, A. N. Core formation on Mars and differentiated asteroids. *Nature* **388**, 854–857 (1997).
- Tonks, W. B. & Melosh, H. J. in *Origin of the Earth* (eds Newsom, N. E. & Jones, J. H.) 151–174 (Oxford Univ. Press, New York, 1990).
- Borg, L. E., Nyquist, L. E., Taylor, L. A., Wiesmann, H. & Shih, C.-Y. Constraints on martian differentiation processes from Rb-Sr and Sm-Nd isotopic analyses of the basaltic shergottite QUE 94201. *Geochim. Cosmochim. Acta* **61**, 4915–4931 (1997).
- Jagoutz, E., Jotter, R. & Dreibus, G. Evolution of six SNC meteorites with anomalous neodymium-142. *Meteorit. Planet. Sci.* **35** (Suppl.), abstr. A83 (2000).
- Borg, L. E., Nyquist, L. E., Wiesmann, H., Shih, C.-Y. & Reese, Y. The age of Dar al Gani 476 and the differentiation history of the martian meteorites inferred from their radiogenic isotopic systematics. *Geochim. Cosmochim. Acta* **67**, 3519–3536 (2003).
- Jagoutz, E., Dreibus, G. & Jotter, R. New ^{142}Nd data on SNC meteorites. *Geochim. Cosmochim. Acta* **67** (Suppl. 1), A184 (2003).
- Boyet, M. & Carlson, R. W. ^{142}Nd evidence for early (>4.53 Ga) global differentiation of the silicate Earth. *Science* **309**, 576–581 (2005).
- Andreasen, R. & Sharma, M. Solar nebula heterogeneity in p-process samarium and neodymium isotopes. *Science* **314**, 806–809 (2006).
- Boyet, M. & Carlson, R. W. A new geochemical model for the Earth's mantle inferred from ^{146}Sm - ^{142}Nd systematics. *Earth Planet. Sci. Lett.* **250**, 254–268 (2006).
- Rehkämper, M., Gärtner, M., Galer, S. J. G. & Goldstein, S. L. Separation of Ce from other rare-earth elements with application to Sm-Nd and La-Ce chronometry. *Chem. Geol.* **129**, 201–208 (1996).
- Shih, C.-Y. *et al.* Chronology and petrogenesis of young achondrites, Shergotty, Zagami, and ALHA 77005: late magmatism on a geologically active planet. *Geochim. Cosmochim. Acta* **46**, 2323–2344 (1982).
- Jones, J. H. Isotopic relationships among the shergottites, the nakhlites and Chassigny. *Proc. Lunar Planet. Sci. Conf.* **19**, 465–474 (1989).
- Borg, L. E., Nyquist, L. E., Wiesmann, H. & Reese, Y. Constraints on the petrogenesis of martian meteorites from the Rb-Sr and Sm-Nd isotopic systematics of the Iherzolitic shergottites ALHA77005 and LEW88516. *Geochim. Cosmochim. Acta* **66**, 2037–2053 (2002).
- Longhi, J. Complex magmatic processes on Mars: inferences from SNC meteorites. *Proc. Lunar Planet. Sci. Conf.* **21**, 695–709 (1991).
- Blichert-Toft, J., Gleason, J. D., Telouk, P. & Albarède, F. The Lu-Hf isotope geochemistry of shergottites and the evolution of the martian mantle-crust system. *Earth Planet. Sci. Lett.* **173**, 25–39 (1999).
- Brandon, A. D., Walker, R. J., Morgan, J. W. & Goles, G. G. Re-Os isotopic evidence for early differentiation of the martian mantle. *Geochim. Cosmochim. Acta* **64**, 4083–4095 (2000).
- Shih, C.-Y., Nyquist, L. E., Wiesmann, H. & Barrat, J.-A. Age and petrogenesis of picritic shergottite NWA 1068: Sm-Nd and Rb-Sr isotopic studies. *Lunar Planet. Sci.* **34**, abstr. 1439 (2003).
- Warren, P. H. & Wasson, J. T. The origin of KREEP. *Rev. Geophys. Space Phys.* **17**, 73–88 (1979).
- Zhang, M. H. G., Luhmann, J. G., Bougher, S. W. & Nagy, A. F. The ancient oxygen exosphere of Mars: implications for atmosphere evolution. *J. Geophys. Res.* **98**, 10915–10923 (1993).
- Caro, G., Bourdon, B., Birck, J.-L. & Moorbath, S. High-precision $^{142}\text{Nd}/^{144}\text{Nd}$ measurements in terrestrial rocks: constraints on the early differentiation of the Earth's mantle. *Geochim. Cosmochim. Acta* **70**, 164–191 (2006).
- Nyquist, L. E. *et al.* Ages and geologic histories of martian meteorites. *Space Sci. Rev.* **96**, 105–164 (2001).
- Brandon, A. D., Nyquist, L. E., Shih, C.-Y. & Wiesmann, H. Rb-Sr and Sm-Nd isotopic systematics of shergottite NWA 856: crystallization age and implications for alteration of hot desert SNC meteorites. *Lunar Planet. Sci.* **35**, abstr. 1931 [CD-ROM] (2004).
- Shih, C.-Y., Nyquist, L. E. & Reese, Y. Rb-Sr and Sm-Nd isotopic studies of martian depleted shergottites SaU094/005. *Lunar Planet. Sci.* **38**, abstr. 1745 [CD-ROM] (2007).
- Jagoutz, E. & Wänke, H. Sr and Nd isotopic systematics of Shergotty meteorite. *Geochim. Cosmochim. Acta* **50**, 939–953 (1986).
- Bouvier, A., Blichert-Toft, J., Vervoort, J. D. & Albarède, F. The age of SNC meteorites and the antiquity of the martian surface. *Earth Planet. Sci. Lett.* **240**, 221–233 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Draper for a review. We thank the Smithsonian Institution, the American Museum of Natural History, the Natural History Museum in Bern, Switzerland, and the NASA Antarctic Meteorite Collection for providing samples for this study. We also thank Y. Reese and C.-Y. Shih for their analytical support and K. Rankenburg for his help. This work was supported by the Lunar and Planetary Institute, NASA Cosmochemistry and Origins of Solar Systems grants.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to V.D. (debaille@lpi.usra.edu).

Coherent zero-state and π -state in an exciton–polariton condensate array

C. W. Lai^{1,2,3}, N. Y. Kim^{1,2}, S. Utsunomiya^{3,4}, G. Roumpos¹, H. Deng¹, M. D. Fraser¹, T. Byrnes^{2,3}, P. Recher^{1,2}, N. Kumada⁴, T. Fujisawa⁴ & Y. Yamamoto^{1,3}

The effect of quantum statistics in quantum gases and liquids results in observable collective properties among many-particle systems. One prime example is Bose–Einstein condensation, whose onset in a quantum liquid leads to phenomena such as superfluidity and superconductivity. A Bose–Einstein condensate is generally defined as a macroscopic occupation of a single-particle quantum state, a phenomenon technically referred to as off-diagonal long-range order due to non-vanishing off-diagonal components of the single-particle density matrix^{1–3}. The wavefunction of the condensate is an order parameter whose phase is essential in characterizing the coherence and superfluid phenomena^{4–11}. The long-range spatial coherence leads to the existence of phase-locked multiple condensates in an array of superfluid helium¹², superconducting Josephson junctions^{13–15} or atomic Bose–Einstein condensates^{15–18}. Under certain circumstances, a quantum phase difference of π is predicted to develop among weakly coupled Josephson junctions¹⁹. Such a meta-stable π -state was discovered in a weak link of superfluid ³He, which is characterized by a ‘p-wave’ order parameter²⁰. The possible existence of such a π -state in weakly coupled atomic Bose–Einstein condensates has also been proposed²¹, but remains undiscovered. Here we report the observation of spontaneous build-up of in-phase (‘zero-state’) and antiphase (‘ π -state’) ‘superfluid’ states in a solid-state system; an array of exciton–polariton condensates connected by weak periodic potential barriers within a semiconductor microcavity. These in-phase and antiphase states reflect the band structure of the one-dimensional polariton array and the dynamic characteristics of metastable exciton–polariton condensates.

The strong coupling between quantum well excitons and microcavity photons allows studies of cavity quantum electrodynamic effects in a solid-state system. The reversible energy exchange between the microcavity photons and quantum well excitons leads to the anti-crossing of eigenstates, known as the upper and lower exciton–polaritons (UPs and LPs)²². Unlike exciton–polaritons in bulk semiconductors, cavity exciton–polaritons involve no transport and are fully characterized by the in-plane wavenumber $k_{||}$, enabling a direct measurement of the exciton–polariton dispersion relation and momentum distribution of particles through angularly resolved spectroscopy²³. The exciton–polaritons have been shown to condense in momentum space and exhibit spontaneous build-up of macroscopic spatial and temporal coherence^{24–26}. Using Young’s double-slit interference experiment, we measured a spatial coherence length of the condensate up to $\sim 20\ \mu\text{m}$, limited by the pumping spot size (see Supplementary Information).

In our experiments, an array of one-dimensional cigar-shaped exciton–polaritons is created by the deposition of periodic strips of a metallic thin film on the top surface of the microcavity structure, as

shown in Fig. 1a (see also Methods). The GaAs-based microcavity structure contains 12 quantum wells, exhibiting a vacuum Rabi splitting of $\sim 14\ \text{meV}$. For the given microcavity structure under the metallic layer, the cavity resonance energy E_C (for $k_{||} = 0$) is expected to increase by $\sim 400\ \mu\text{eV}$ according to the transfer matrix method²⁷

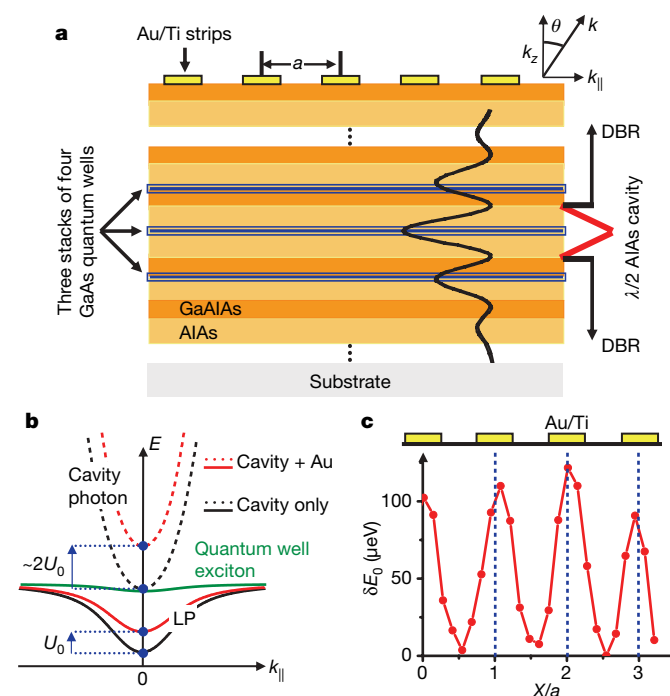


Figure 1 | Formation of an exciton–polariton array. **a**, A schematic of a cavity polariton array formed by depositing periodic thin metallic strips (Au/Ti) on top of a microcavity structure. The $1.4\text{-}\mu\text{m}$ -wide strips are equally spaced with a pitch distance $a = 2.8\ \mu\text{m}$. The microcavity structure consists of a $\lambda/2$ AlAs cavity (indicated by red lines) sandwiched by two distributed Bragg reflectors with alternating GaAlAs/AlAs $\lambda/4$ layers, where λ is the cavity resonance wavelength (varying around the quantum well exciton resonance, $\sim 776\ \text{nm}$ with tapering). Three stacks of four GaAs quantum wells are positioned at the central three antinodes of the microcavity photon field (the schematic black oscillatory curve). The distributed Bragg reflectors (DBR) consist of repeated GaAlAs/AlAs layers and are symbolized by the two short dotted vertical lines. **b**, Dispersion curves of the cavity photon mode (dashed lines) and LP (solid lines). The black dashed lines represent the uncoupled cavity photon mode and quantum-well heavy-hole exciton state, which are in resonance ($E_C = E_X$) at $k_{||} = 0$ in this case. The addition of the metallic layer shifts the LP energy (red solid line) by $U_0 \approx 200\ \mu\text{eV}$. **c**, Spatial LP energy modulation ΔE_0 measured by scanning a pinhole across the lattice at the first conjugate real-image plane (see Fig. 2 and Methods).

¹E. L. Ginzton Laboratory, Stanford University, Stanford, California 94305, USA. ²Institute of Industrial Science, University of Tokyo, Meguro-ku, Tokyo 153-8505, Japan. ³National Institute of Informatics, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan. ⁴NTT basic research laboratories, NTT Corporation, Atugi, Kanagawa 243-0198, Japan.

(see Methods). When the cavity photon (E_C) is near resonance with the quantum well exciton (E_X) at $k_{||} = 0$, the shift of the lower-polariton energy induced by the metallic layer is $\sim 200 \mu\text{eV}$, approximately half of the cavity photon resonance shift (Fig. 1b). Therefore, the LPs are expected to be trapped in the gap region where the LP energy is lower. The measured spatial modulation of LP energy is $\sim 100 \mu\text{eV}$, which is less than the theoretical prediction owing to the diffraction-limited spatial resolution ($\sim 2 \mu\text{m}$) of our optical detection system (Fig. 1c). We confirm by an independent measurement that the LP energy in a bare microcavity is $\sim 200 \mu\text{eV}$ lower than the LP energy under a uniform metallic layer of similar thickness for $E_C \approx E_X$.

The microcavity is excited by a mode-locked Ti:Sapphire laser with a ~ 2.5 ps pulse near the quantum well exciton resonance at an incident angle of 60° , corresponding to an in-plane wavenumber $k_{||} \approx 7 \times 10^4 \text{ cm}^{-1}$ in air. The large $k_{||}$ of the pump ensures that the coherence of the pump laser is lost by multiple phonon emissions before the polaritons scatter into $k_{||} \approx 0$ states. The LP distributions in coordinate space (near-field) and momentum space (far-field) are measured by micro-photoluminescence spectroscopy and imaging, using the set-up shown in Fig. 2a (see Methods). The near-field images illustrate the spatial LP distribution, while the far-field images reveal the in-plane momentum LP distribution. The spectroscopy in momentum space (that is, angularly resolved spectroscopy) provides a direct measurement of the LP dispersion relation and effective mass.

For a system of two condensates, the interference fringes can be observed even in the absence of a locked relative phase in a single shot measurement⁷. In the case of an array of condensates, however, interference fringes appear only if the relative phases between consecutive condensates are locked²⁸. The high degree of interference fringe visibility in momentum space can thus be used as an indication of a long-range spatial coherence across the whole array of multiple condensates. Such an interference measurement is commonly performed in atomic Bose–Einstein condensates by releasing condensates in an optical lattice.

Considering a periodic and coherent array of condensates aligned along the x (where x is X or Y , see below) axis, we can approximate the order parameter in momentum space as²⁸:

$$\begin{aligned}\psi(k_x) &= \psi_0(k_x) \sum_{n=0, \pm 1, \dots, \pm N_M} e^{i n (k_x a + \phi)} \\ &= \psi_0(k_x) \frac{\sin(N k_x a/2)}{\sin(k_x a/2)} \Big|_{\phi=0}\end{aligned}\quad (1)$$

where n labels the condensate array element, $N = 2N_M + 1$ is the total number of periodic array elements, a is the pitch distance between two neighbouring elements, and $\psi_0(k_x)$ is the momentum-space wavefunction of an individual element. The overall momentum distribution, $\rho(k_x) = |\psi(k_x)|^2$, reflects the coherence properties and nature of the condensate arrays through distinctive interference patterns. In the presence of the lattice, the momentum distribution displays narrow peaks at $k_x = m \times 2\pi/a$ (where m is an integer) with an envelope function given by $\rho_0(k_x) = |\psi_0(k_x)|^2$.

If all the elements are locked in-phase ($\phi = 0$), at a LP wavelength $\lambda = 780 \text{ nm}$ and a grating pitch distance $a = 2.8 \mu\text{m}$, a Fraunhofer diffraction pattern has a central peak at $\theta = 0^\circ$ and two side lobes at $\theta = \sin^{-1}(\lambda/a) \approx \pm 16^\circ$ ($k_{||} = \pm 2\pi/a$; $m = \pm 1$ first-order diffraction). The LP emissions from an exciton–polariton condensate array are expected to display such an interference pattern in momentum space, just like an atomic Bose–Einstein condensate experiment^{15–18}.

The observed LP distribution in both coordinate and momentum space of the condensate array are shown in Fig. 2b and c. Below the condensation threshold (pump power $P = 20 \text{ mW}$), the near-field image reveals the cigar-shaped LP emissions through the $1.4\text{-}\mu\text{m}$ -wide gaps between the metallic strips. The corresponding LP momentum distribution is broad ($\Delta\theta_{\text{FWHM}} \approx 36^\circ$) and isotropic, independent of the periodic intensity modulation and the elliptically shaped pumping spot. The result indicates that there is negligible phase coherence among different cigar-shaped LPs. With an increasing pumping rate, two strong and two weak side lobes emerge at

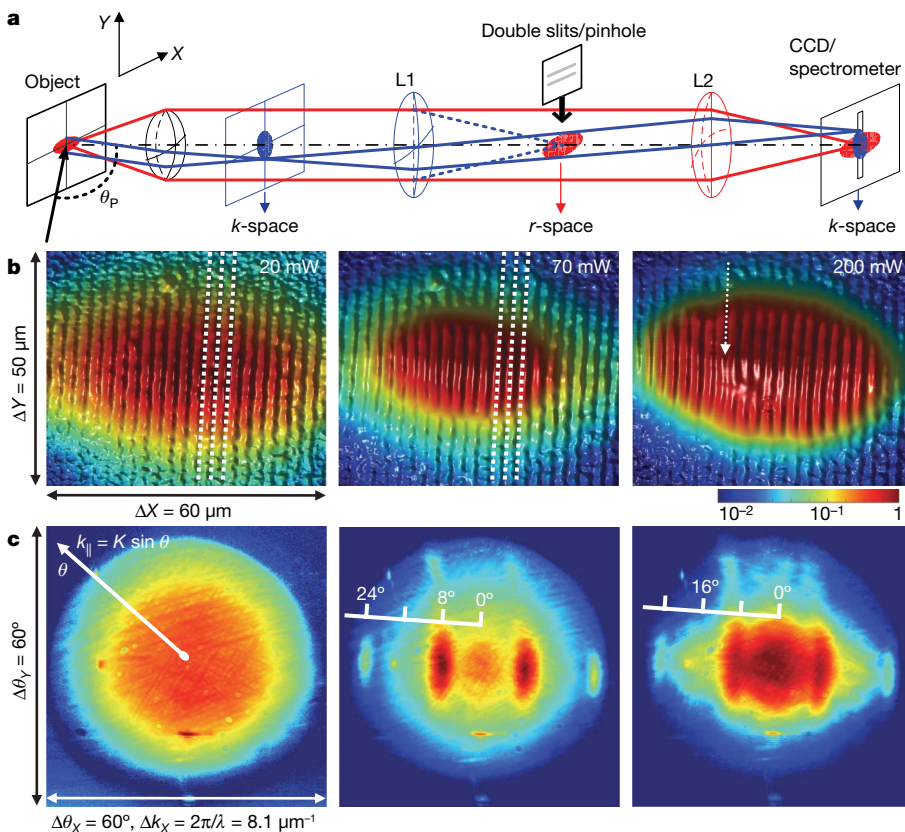


Figure 2 | Imaging and spectroscopy of exciton–polariton distribution in coordinate and momentum space. **a**, Schematic of the micro-photoluminescence set-up for imaging and spectroscopy in coordinate and momentum space (see Methods). $\theta_p \approx 60^\circ$ is the incident angle of the pump beam. **b**, Near-field images showing the LP distribution across a polariton array in coordinate space under pumping powers of 20, 70 and 200 mW (left to right) at $T = 20 \text{ K}$. Here the condensation threshold pumping power is $\sim 45 \text{ mW}$ (intensity $\sim 1,600 \text{ W cm}^{-2}$). The white dashed lines indicate selected locations of the $1.4\text{-}\mu\text{m}$ -wide metallic strips. The colour scale is the normalized LP emission intensity. **c**, Corresponding far-field images showing the LP distribution in momentum space. When passing through the threshold, lobes at $\pm 8^\circ$ and $\pm 24^\circ$ emerge out of the isotropic background of the thermal polariton gas, which is indicative of the π -state. The strong central lobe at 0° and two weak lobes at $\pm 16^\circ$ appear at higher pumping rates, indicative of the zero-state.

$\theta \approx \pm 8^\circ$ and $\pm 24^\circ$, which correspond to $k_{||} = \pm \pi/a$ and $k_{||} = \pm 3\pi/2a$, respectively. The diffraction pattern suggests that the phase difference of adjacent condensates in the array is locked exactly to π (see $\phi = \pi$ in equation (1)). Moreover, the corresponding near-field image (in Fig. 2b; 70 mW) reveals that strong LP emissions are generated from under the metallic strips rather than the gaps. The contrast is better observed near the boundary of the central condensate and the outside thermal LP emissions, which come through the gaps. The dominant LP emission through the metallic layers despite the lower transmission indicates that LP condensates are strongly localized under the metallic strips. With a further increasing pumping rate, the intensity of the central peak near $\theta = 0^\circ$ with weak side lobes at $\theta \approx \pm 16^\circ$ (Fig. 2c) gradually surpasses the peaks at $\theta \approx \pm 8^\circ$ and $\pm 24^\circ$ as shown in Fig. 2c. The corresponding near-field image of the condensate also recovers a standard spatial modulation (near the central area indicated by the white arrow in Fig. 2b).

These two distinct interference patterns suggest the transition from an antiphase state to an in-phase state. We refer to the in-phase state as the ‘zero-state’ and the antiphase state as the ‘ π -state’ on the basis of their relative phase differences between adjacent elements.

Both the zero-state and π -state can be observed for an exciton–polariton condensate array consisting of more than 30 strips of condensates (total array dimension $\sim 100 \mu\text{m}$). These collective states are manifestations of the long-range spatial coherence as well as of phase-locking across the array. Next, we examine the phase-locking mechanism for the zero-state and π -state through the LP dispersion relation.

In Fig. 3a, we show the energy versus in-plane momentum spectra for an LP array near resonance $E_C \approx E_X$ and below threshold. From the dispersion curve, we deduce the LP kinetic energy at $k_{||} = \pm G_0/2 = \pi/a$ (half of the primitive reciprocal lattice vector), $E_k = \hbar^2(G_0/2)^2/2m^* \approx 500 \mu\text{eV}$, where the LP effective mass $m^* = 9 \times 10^{-5} m_e$ (where m_e is the free electron mass). The kinetic energy E_k is larger than the spatial LP energy modulation, $U_0 \approx 200 \mu\text{eV}$. Therefore, we expect the exciton–polariton condensate to be only weakly perturbed by the periodic potential and the coherence is maintained across the array. In such a large tunnelling limit, we neglect the polariton–polariton interaction and deduce the band structure of the polariton array by assuming a ‘nearly free polariton’ in the presence of the periodic square-well potential. Given a one-dimensional periodic potential $U_0(X)$ with a lattice constant a , the band structure can be obtained using the standard Bloch wavefunction formalism. Only those polaritons with in-plane momentum close to the Bragg condition ($k_{||} = m G_0/2$) are subject to backscattering, owing to the periodic potential, and gaps appear at the Bragg planes where standing waves are formed. Similar to the standard extended-zone scheme²⁹, the band structure can be constructed by starting with an original parabola and displaced parabolas $E_{LP}(k_{||} \pm m G_0)$, as shown in Fig. 3b. Here the size of the circles represents the expected relative LP emission intensity from each band (Methods). We observe this in the LP energy versus in-plane momentum with a below-threshold pumping rate (Fig. 3a). The first bandgap, between the first and second bands near the zone boundaries (Bragg planes), is of the order of the barrier potential, $|U_0|$. This small gap is masked in the measured spectra by the large inhomogeneous broadening of the LP emission lines (linewidth $\sim 500 \mu\text{eV}$).

In Fig. 3c, we show the energy versus in-plane momentum of a polariton condensate array above threshold and $E_C - E_X \approx +6 \text{ meV}$. Above threshold, LP emissions occur in two states with an energy difference of about 1 meV. These two states with emission peaks at $k_{||} = 0, \pm G_0$ and at $k_{||} = \pm G_0/2, \pm 3G_0/2$ correspond to the zero-state and π -state, respectively.

On referring to the multi-valley band structure (Fig. 3b), the dynamic condensation process of the exciton–polariton array becomes more transparent. Without the spatial modulation, the quasi-stationary state of the LP condensate is the lowest-energy state at $k_{||} = 0$. Here the bottom of the LP dispersion parabola serves as a

‘trap’ of polaritons in momentum space. When the spatial potential modulation is introduced, a meta-stable dynamic condensate can occur near the bottom of the second band (point A in Fig. 3b). At these points, polaritons experience a relaxation bottleneck, resulting in metastable condensates. Eventually, the LP system will relax to the lowest-energy in-phase state (point C in Fig. 3b) as the pumping rate is increased.

To understand the spatial distributions of LPs in the array (near-field images shown in Fig. 2b), we consider the Bloch wavefunctions. The Bloch wavefunctions for these eigenstates exhibit not only opposite relative phases between adjacent elements in the array but also different characteristics. Schematic Bloch wavefunctions for selected eigenstates labelled as A, B and C in Fig. 3b are shown in Fig. 3d. The zero-state at point C corresponds to the ‘s-like’ state with maximal amplitude in potential wells and identical phase between adjacent elements across the array. The wavefunctions of A and B at $k_{||} = G_0/2$ both exhibit a relative π -phase difference between adjacent elements. The lower-energy state at point B corresponds to anti-bonding of ‘s-states’, while the higher-energy state at point A corresponds to bonding of ‘p-states’. Like the atomic p-state, the LP density vanishes at the centre of the potential well for this π -state. This is fully confirmed by the near-field imaging shown in Fig. 2b. We note that under the weak potential modulation, the Bloch wavefunctions are superpositions of forward-and-backward travelling plane waves at the zone boundary and have a relatively broad distribution in space compared to the standard tight-binding model of a normal condensed-matter system.

The metastability of the π -state is a unique property for dynamic polariton condensates. Under strong pumping rates, the zero-state dominates eventually, owing to enhanced cooling by stimulated polariton–polariton scattering at high densities. A typical evolution

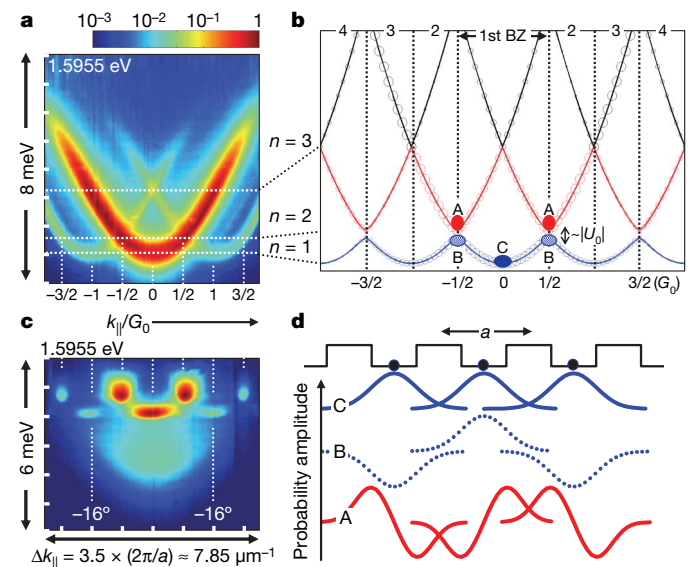


Figure 3 | Band structure and ‘superfluid’ states in an exciton–polariton condensate array. **a**, Time-integrated energy versus in-plane momentum (E versus $k_{||}$) near resonance $E_C \approx E_X$ below threshold ($P = 10 \text{ mW}$) at $T \approx 7 \text{ K}$. The central bright parabola corresponds to the dispersion curve for the polaritons in the absence of a periodic potential. Two additional parabolas displaced by $\pm G_0 = \pm 2\pi/a$ cross the central dispersion curve at $k_{||} = \pm G_0/2$. **b**, Extended-zone scheme of the band structure. Anti-crossing occurs at the boundary between the first and second Brillouin zones (BZ). The size of the open circles represents the expected relative emission intensity (log scale) of each energy band. Condensation occurs at the valleys labelled by solid red (point A) and blue (point C) circles. **c**, E versus $k_{||}$ at blue detuning $E_C - E_X \approx 6 \text{ meV}$ above threshold ($P = 40 \text{ mW}$). The lobes at $k_{||} = 0, \pm G_0$ correspond to the in-phase zero-state, whereas the lobes at $k_{||} = \pm G_0/2, \pm 3G_0/2$ correspond to the antiphase π -state. **d**, Schematic Bloch-wave functions for states labelled as A, B or C in **b**.

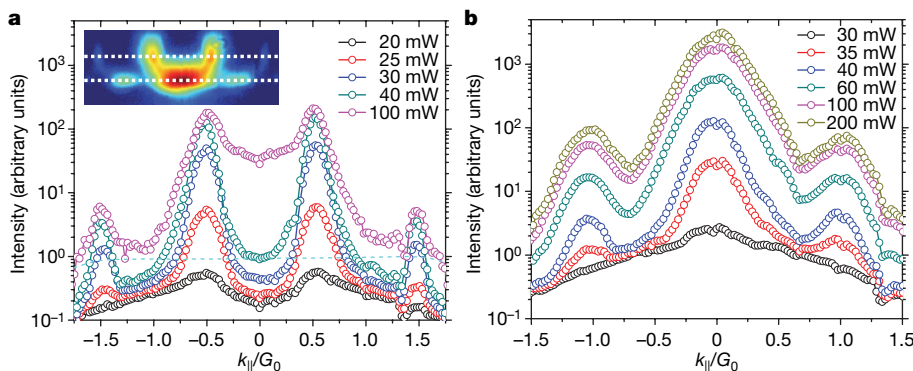


Figure 4 | Evolution of the 'zero-state' and 'pi-state'. **a**, Profiles of the π -state as a function of in-plane momentum taken at a cross-section in the dispersion curve (exemplified by the white dotted lines in the inset) for an exciton–polariton condensate array for various pumping rates. The inset shows the energy versus in-plane momentum spectra (axes and colour scale same as Fig. 3c, but only the top 2 meV is shown in the inset) for $P = 100$ mW, where relaxation from the π -state to zero-state is visible. The blue dashed line displays the uniform collection efficiency of the optical system up to $|k_{\parallel}| = 3G_0/2$. **b**, Profiles of the zero-state. The intensity of the zero-state surpasses that of the π -state at around $P = 45$ mW.

from the metastable π -state to the stationary zero-state is illustrated by the LP momentum distribution profiles for various pumping rates in Fig. 4. At a high pumping rate, the LP emission from the zero-state surpasses that from the π -state by more than an order of magnitude. The transition between these dynamic condensates depends on the coupling strength between adjacent elements in the array (see Supplementary Information).

Thus we observed a metastable condensate in a π -state (anti-bonding of p-states) as well as a stable condensate in a zero-state (bonding of s-states) in an exciton–polariton array. Such a coherent linear superposition state formed over many condensates in periodic sites is referred to as a 'superfluid' state. However, some common superfluidity properties such as the frictionless flow over a long distance beyond the pumping region are not necessarily meant. Our experiments constitute a first step towards the experimental study of a dynamical Bose–Hubbard model in a solid-state system. The polariton system allows the direct observation of the spatial and momentum distributions as well as the dispersion characteristics of the condensates. For instance, full control of the on-site interaction, barrier potential and occupation number may allow us to study the dynamical quantum phase transition properties of a Mott insulator.

METHODS SUMMARY

The periodic spatial modulation of LP energy ~ 200 μ eV is induced by the periodic thin metallic strips (a ~ 20 -nm-thick Au / 3-nm-thick Ti layer) with a pitch distance $a = 2.8$ μ m deposited on top of the microcavity. The relative transmission through the top distributed Bragg reflectors (DBR) with and without the metallic layer is ~ 40 –50%.

The imaging and spectroscopy of LPs in the near-field (coordinate space) and far-field (momentum space) are performed by positioning a charged-coupled-device (CCD) camera or an imaging spectrometer in a plane conjugate to the real-image or Fourier (back focal) plane of the objective. The collection efficiency is approximately uniform to a full angle of $\sim 60^\circ$.

More details about the preparation of the metallic grating layer and its effects on the far-field interference pattern, the experimental settings, including the determination of the in-plane wavevector in momentum space, and the expected LP diffraction patterns and intensities of the zero-state and π -state are given in the Methods section and Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 May; accepted 20 September 2007.

1. Penrose, O. & Onsager, L. Bose-Einstein condensation and liquid helium. *Phys. Rev.* **104**, 576–584 (1956).
2. Beliaev, S. T. Application of the methods of quantum field theory to a system of bosons. *J. Exp. Theor. Phys.* **34**, 417–432 (1958).
3. Yang, C. N. Concept of off-diagonal long-range order and the quantum phases of liquid He and of superconductors. *Rev. Mod. Phys.* **34**, 694–704 (1962).
4. Tilley, D. R. & Tilley, J. *Superfluidity and Superconductivity* Chs 1–3, and 7 (Adam Hilger, New York, 1990).
5. Pitaevskii, L. P. & Stringari, S. *Bose-Einstein Condensation* (Clarendon, Oxford, 2003).

6. Leggett, A. J. *Quantum Liquids: Bose Condensation and Cooper Pairing in Condensed-matter Systems* Chs 1–4 (Oxford Univ. Press, Oxford, 2006).
7. Andrews, M. R. *et al.* Observation of interference between two Bose condensates. *Science* **275**, 637–641 (1997).
8. Hagley, E. W. *et al.* Measurement of the coherence of a Bose-Einstein condensate. *Phys. Rev. Lett.* **83**, 3112–3115 (1999).
9. Bloch, I., Hansch, T. W. & Esslinger, T. Measurement of the spatial coherence of a trapped Bose gas at the phase transition. *Nature* **403**, 166–170 (2000).
10. Castin, Y. & Dalibard, J. Relative phase of two Bose-Einstein condensates. *Phys. Rev. A* **55**, 4330–4337 (1997).
11. Naraschewski, M. & Glauber, R. J. Spatial coherence and density correlations of trapped Bose gases. *Phys. Rev. A* **59**, 4595–4607 (1999).
12. Davis, J. C. & Packard, R. E. Superfluid 3 He Josephson weak links. *Rev. Mod. Phys.* **74**, 741–773 (2002).
13. Hansen, J. B. & Lindelof, P. E. Static and dynamic interactions between Josephson junctions. *Rev. Mod. Phys.* **56**, 431–459 (1984).
14. Hadley, P. M., Beasley, R. & Wiesenfeld, K. Phase locking of Josephson-junction series arrays. *Phys. Rev. B* **38**, 8712–8719 (1988).
15. Cataliotti, F. S. *et al.* Josephson junction arrays with Bose-Einstein condensates. *Science* **293**, 843–846 (2001).
16. Anderson, B. P. & Kasevich, M. A. Macroscopic quantum interference from atomic tunnel arrays. *Science* **282**, 1686–1689 (1998).
17. Orzel, C. *et al.* Squeezed states in a Bose-Einstein condensate. *Science* **291**, 2386–2389 (2001).
18. Greiner, M. *et al.* Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms. *Nature* **415**, 39–44 (2002).
19. Bulaevskii, L. N., Kuzii, V. V. & Sobyannin, A. A. Superconducting system with weak coupling to current in ground state. *JETP Lett.* **25**, 290–294 (1977).
20. Backhaus, S. *et al.* Discovery of a metastable pi-state in a superfluid 3 He weak link. *Nature* **392**, 687–690 (1998).
21. Smerzi, A. *et al.* Quantum coherent atomic tunneling between two trapped Bose-Einstein condensates. *Phys. Rev. Lett.* **79**, 4950–4953 (1997).
22. Weisbuch, C. *et al.* Observation of the coupled exciton-photon mode splitting in a semiconductor quantum microcavity. *Phys. Rev. Lett.* **69**, 3314–3317 (1992).
23. Houdre, R. *et al.* Measurement of cavity-polariton dispersion curve from angle-resolved photoluminescence experiments. *Phys. Rev. Lett.* **73**, 2043–2046 (1994).
24. Deng, H. *et al.* Polariton lasing versus photon lasing in a semiconductor microcavity. *Proc. Natl Acad. Sci. USA* **100**, 15318–15323 (2003).
25. Deng, H. *et al.* Quantum degenerate exciton-polaritons in thermal equilibrium. *Phys. Rev. Lett.* **97**, 146402 (2006).
26. Kasprzak, J. *et al.* Bose-Einstein condensation of exciton polaritons. *Nature* **443**, 409–414 (2006).
27. Yeh, P. *Optical Waves in Layered Media* Ch. 5 (John Wiley, Hoboken, 2005).
28. Pedri, P. *et al.* Expansion of a coherent array of Bose-Einstein condensates. *Phys. Rev. Lett.* **87**, 220401 (2001).
29. Madelung, O. *Introduction to Solid-state Theory* Ch. 2 (Springer, Berlin, 1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the JST/SORST programme and by Special Coordination Funds for Promoting Science and Technology in Japan. The high-quality microcavity cavity sample is courtesy of G. S. Solomon, R. Hey, K. Ploog and A. Forchel. We thank T. Maruyama for support and S. Sasaki for the device fabrication. C.W.L. thanks Y. R. Shen for comments and discussions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.W.L. (cwlai@stanford.edu) or Y.Y. (yyamamoto@stanford.edu).

METHODS

Exciton–polariton array. The thin metallic film deposited on top of the microcavity modifies the boundary condition of the top DBR structure. For a bare microcavity structure, the resonant photon mode exhibits an anti-node at the semiconductor/free-space interface. In the presence of a metallic layer, the anti-node shifts towards the centre of the cavity, owing to suppression of the electromagnetic field inside the metallic layer. As a result, the energy of the cavity photon resonance increases (blue shift). For a 20-nm-thick Au/3-nm-thick Ti layer on top of the microcavity structure, the cavity resonance energy increases by ~ 400 μeV according to the calculation by the transfer matrix method²⁷. The associated LP energy shift is of the order of ~ 200 μeV near-resonance $E_C \approx E_X$, approximately half of the cavity resonance shift. This is consistent with the experimentally measured energy shift under a uniform metallic layer of similar thickness.

The relative transmittance through the top DBR with and without the metallic layer is ~ 40 – 50% , so the optical pumping and LP emission through the periodic metallic strips are spatially modulated. The initial spatial modulation of LP generation by the metallic strips is mostly washed out by internal diffusion of polaritons through the cooling process. However, the detected LP emission is subject to intensity modulation by the periodic metallic strips. This effect must be separately considered for understanding the far-field interference pattern from the spontaneously formed ‘zero-state’ and ‘ π -state’ under the built-in periodic potential.

Experimental techniques. The exciton–polaritons are excited by a p-polarized Ti:sapphire laser of 2.5 ps duration (0.5 meV spectral width) at a 76 MHz repetition rate. The pulses are focused to an elliptical spot ($\sim 60 \mu\text{m} \times 30 \mu\text{m}$ or $\sim 100 \mu\text{m} \times 50 \mu\text{m}$) onto the sample at an oblique incidence angle of $\sim 60^\circ$ ($k_{\parallel} \approx 7 \times 10^4 \text{ cm}^{-1}$ in air). The elliptical pumping spot covers approximately 20–30 periodic elements.

LPs are trapped in the cavity mostly until they relax to states of lower energy and in-plane momentum. The emitted luminescence is collected by a micro-photoluminescence set-up composed of a $50\times$ objective lens (effective focal length, $f_{\text{obj}} = 4 \text{ mm}$ and a numerical aperture, $\text{NA} = 0.55$, corresponding to a full collection angle of $\sim 67^\circ$ in air) and a pair of relay lenses of focal length $f_L = 10 \text{ cm}$ (L1, which is removable, and L2) as shown in Fig. 2. The collection efficiency is approximately uniform up to a full angle $\Delta\theta \approx 60^\circ$ ($\Delta k_{\parallel} \approx 2\pi/\lambda$).

The imaging and spectroscopy of LPs in the near-field (coordinate space) and far-field (momentum space) are performed by positioning a charged-coupled-device (CCD) camera or an imaging spectrometer in a plane conjugate to the real-image or Fourier (back focal) plane of the objective. The optical axis of the collection optical system is perpendicular to the surface of the sample. For imaging and spectroscopy in coordinate space (near-field), the repositionable lens L1 is removed. The Fourier (back focal) plane of the sample located in the object plane, with coordinates X and Y , is thus imaged onto to a CCD camera directly or through an imaging spectrometer. This arrangement allows determination of the LP distribution in momentum space (k_X, k_Y) (far-field imaging) as well as the LP dispersion relation $E_{\text{LP}}(k_{\parallel})$. Coordinate space is correlated to momentum space by a Fourier transform: $(X, Y) = (u/M, v/M) \leftrightarrow (k_X, k_Y) = (k \times u/f_{\text{obj}}, k \times v/f_{\text{obj}})$, where M is the magnification ($M = f_L/f_{\text{obj}} \approx 25$), (u, v) are coordinates of the near-field or far-field images, and $k_{\parallel} = k_{X,Y} = k \sin \theta = (2\pi/\lambda) \sin \theta$ in air. Near-field and far-field images are measured with a CCD with a pixel size of $\sim 8 \times 8 \mu\text{m}^2$, giving a diffraction-limited spatial resolution of $\sim 2 \mu\text{m}$ and a pixel-size-limited angular resolution of $\delta\theta = 0.11^\circ$. The energy versus in-plane momentum relation of LPs and LP arrays are determined by spectroscopy in momentum space (angle-resolved spectroscopy). The spectroscopic system is a 75-cm spectrometer equipped with an $1,800 \text{ g mm}^{-1}$ grating and a liquid-nitrogen-cooled CCD with a pixel size of $26 \times 26 \mu\text{m}^2$. The system has an angular resolution of $\delta\theta \approx 0.4^\circ$ and a spectral resolution of $\sim 20 \mu\text{eV}$, limited by the pixel size.

LP emission intensity from the array. The probability amplitude of LPs in the presence of a periodic potential is represented by the Bloch wave function $\psi_{n,k_{\parallel}}$, where k_{\parallel} is the in-plane lattice momentum, and n the index of the band. The expected leakage LP emission intensity is proportional to the overlapping probability density $\rho = \left| \langle \phi_{k_{\parallel}} | \psi_{n,k_{\parallel}} \rangle \right|^2$, where $\phi_{k_{\parallel}} = \exp[ik_{\parallel} \cdot r]$ is the plane wave corresponding to the free-moving polaritons (that is, photons leaking out from the cavity) with in-plane momentum k_{\parallel} . The calculated relative ρ of each energy band is represented by the size of the open circles (log scale) in Fig. 3b.

Electron pockets in the Fermi surface of hole-doped high- T_c superconductors

David LeBoeuf¹, Nicolas Doiron-Leyraud¹, Julien Levallois², R. Daou¹, J.-B. Bonnemaison¹, N. E. Hussey³, L. Balicas⁴, B. J. Ramshaw⁵, Ruixing Liang^{5,6}, D. A. Bonn^{5,6}, W. N. Hardy^{5,6}, S. Adachi⁷, Cyril Proust² & Louis Taillefer^{1,6}

High-temperature superconductivity in copper oxides occurs when the materials are chemically tuned to have a carrier concentration intermediate between their metallic state at high doping and their insulating state at zero doping. The underlying evolution of the electron system in the absence of superconductivity is still unclear, and a question of central importance is whether it involves any intermediate phase with broken symmetry¹. The Fermi surface of the electronic states in the underdoped 'YBCO' materials $\text{YBa}_2\text{Cu}_3\text{O}_y$ and $\text{YBa}_2\text{Cu}_4\text{O}_8$ was recently shown to include small pockets^{2–4}, in contrast with the large cylinder that characterizes the overdoped regime⁵, pointing to a topological change in the Fermi surface. Here we report the observation of a negative Hall resistance in the magnetic-field-induced normal state of $\text{YBa}_2\text{Cu}_3\text{O}_y$ and $\text{YBa}_2\text{Cu}_4\text{O}_8$, which reveals that these pockets are electron-like rather than hole-like. We propose that these electron pockets most probably arise from a reconstruction of the Fermi surface caused by the onset of a density-wave phase, as is thought to occur in the electron-doped copper oxides near the onset of antiferromagnetic order^{6,7}. Comparison with materials of the La_2CuO_4 family that exhibit spin/charge density-wave order^{8–11} suggests that a Fermi surface reconstruction also occurs in those materials, pointing to a generic property of high-transition-temperature (T_c) superconductors.

The Hall effect is a powerful probe of the Fermi surface of a metal because of its sensitivity to the sign of charge carriers, which distinguishes between electrons and holes. In addition, the Hall effect has been the prime transport signature of density-wave order in copper oxides such as $\text{La}_{2-y}\text{Nd}_y\text{Sr}_x\text{CuO}_4$ (Nd-LSCO) (ref. 10) and $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$ (LBCO) (ref. 11). The Hall resistance R_{xy} of LBCO is reproduced in Fig. 1, in which it drops precipitously below a temperature T_{DW} that coincides with the well-established onset of spin/charge density-wave order in this material⁸. The drop leads to a change of sign in R_{xy} , pointing to a reconstruction of the Fermi surface from purely hole-like above T_{DW} to a combination of electron-like and hole-like sheets below T_{DW} . The fact that our high-field measurement of R_{xy} in $\text{YBa}_2\text{Cu}_3\text{O}_y$, a copper oxide material with a different structure and considerably higher purity, cation order and maximal T_c , exhibits a similar behaviour of R_{xy} , as shown in Fig. 1, raises the possibility that Fermi surface reconstruction may be a generic phenomenon in copper oxides, and hence is likely to be essential for a full understanding of high-temperature superconductors.

The Hall resistance was measured in two closely related underdoped copper oxides of the YBCO family: $\text{YBa}_2\text{Cu}_3\text{O}_y$ (Y123), with $y = 6.51$ and $y = 6.67$, and $\text{YBa}_2\text{Cu}_4\text{O}_8$ (Y124). The Y123 samples

have a high degree of oxygen order, with ortho-II and ortho-VIII superstructure, respectively. The Y124 is stoichiometric, with intrinsic oxygen order. With T_c values of 57.5, 66.0 and 80 K, respectively, the three samples have a hole doping per planar copper atom of $p = 0.10, 0.12$ and 0.14 , respectively, that is, they all fall in the underdoped region of the doping phase diagram ($p < 0.16$). (Sample characteristics are given in the Methods Summary.) The current was applied along the a axis of the orthorhombic structure ($\parallel |x||a$), that is, perpendicular to the CuO chains, in magnetic fields applied normal to the CuO_2 planes ($B \parallel |z||c$). (Details of the measurements are given in the Methods Summary.) The Hall coefficient $R_H \equiv t R_{xy}/B$, where t is the sample thickness, is displayed as a function of magnetic field in Fig. 2 and as a function of temperature in Supplementary Fig. 1.

Our central finding is that all three materials have a negative Hall coefficient in the normal state at low temperature. This is displayed in Fig. 3, where a plot of R_H versus T at the highest field reveals a change of sign from $R_H > 0$ above T_0 to $R_H < 0$ below T_0 , with $T_0 = 30, 70$ and 30 K for Y123-II, Y123-VIII and Y124, respectively, with ± 2 K uncertainty. A very similar sign change was reported in ref. 12 in Y123 samples with $T_c = 62$ – 64 K. Because their measurements were limited to moderate fields (below 24 T), these authors attributed the negative R_{xy} to a negative contribution to the Hall conductivity σ_{xy} .

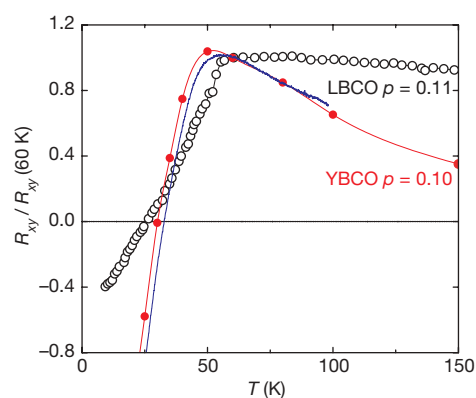


Figure 1 | Hall resistance of LBCO and YBCO. Hall resistance R_{xy} versus T , normalized at 60 K, for LBCO at $p = 0.11$ ($x = 0.11$; black circles from ref. 11) and YBCO at $p = 0.10$. Our data on YBCO were obtained on two different Y123-II samples (with $y = 6.51$), one measured in a continuous temperature sweep at a constant field of 45 T (at the NHMFL; blue curve) and the other measured via field sweeps up to 61 T (at the LNCMP; red circles, taken at 55 T).

¹Département de physique and RQMP, Université de Sherbrooke, Sherbrooke J1K 2R1, Canada. ²Laboratoire National des Champs Magnétiques Pulsés (LNCMP), UMR CNRS-UPS-INSU 5147, Toulouse 31400, France. ³H. H. Wills Physics Laboratory, University of Bristol, Bristol BS8 1TL, UK. ⁴National High Magnetic Field Laboratory (NHMFL), Florida State University, Tallahassee, Florida 32306, USA. ⁵Department of Physics and Astronomy, University of British Columbia, Vancouver V6T 1Z4, Canada. ⁶Canadian Institute for Advanced Research, Toronto M5G 1Z8, Canada. ⁷Superconductivity Research Laboratory, International Superconductivity Technology Center, Shinonome 1-10-13, Koto-ku, Tokyo 135-0062, Japan.

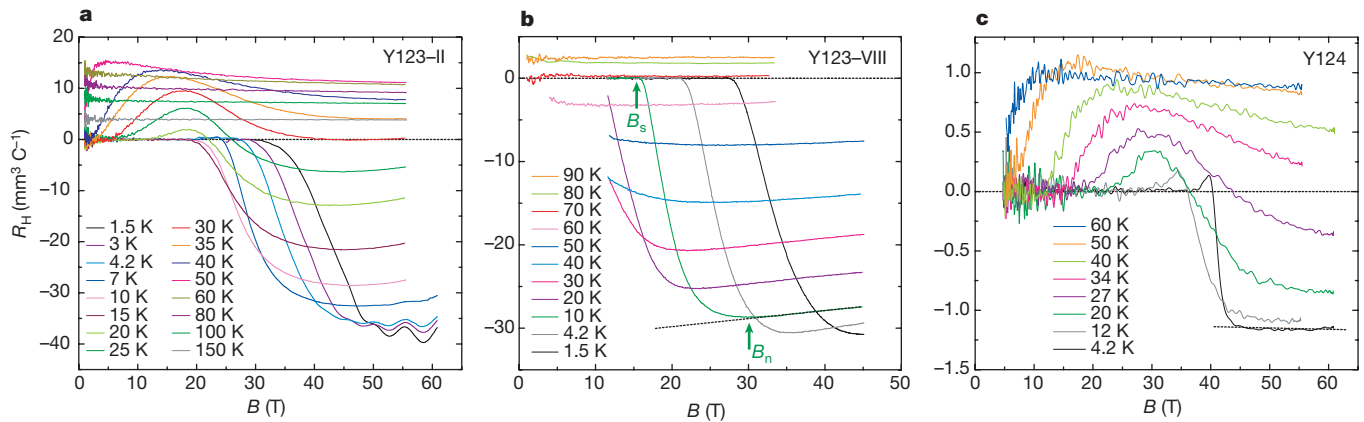


Figure 2 | Hall coefficient versus magnetic field. Hall coefficient $R_H = tR_{xy}/B$ as a function of magnetic field B at the indicated temperatures. **a**, Y123-II ($p = 0.10$); **b**, Y123-VIII ($p = 0.12$); **c**, Y124 ($p = 0.14$). The arrows in **b** indicate the fields B_s and B_n described in the text and defined in the

coming from vortices (flux flow). By going to much higher fields, we can now rule out this interpretation, as discussed in detail in the Supplementary Information, in which the negative R_H is shown to be unambiguously a property of the normal state, the consequence of a drop in $R_H(T)$ that starts below a field-independent temperature T_{max} . The value of T_{max} at the three doping levels studied here is 50, 105 and 60 K, for Y123-II, Y123-VIII and Y124, respectively, with ± 5 K uncertainty (see arrows in Fig. 3).

Three groups have previously detected this drop in low-field measurements of underdoped Y123, with $B < 15$ T, on crystals with $T_c(0) = 60$ –70 K (refs 13–15). Because these earlier studies were limited to high temperatures ($T > T_c(0)$), they failed to reveal that the drop is just the start of a large swing to negative values. By measuring R_{xx} and R_{xy} along both a and b axes, it was shown¹⁵ that the drop in $R_H(T)$ is a property of the planes, not the chains. From the perfect linearity of R_{xy} versus B it was also concluded that the drop is not due to flux flow¹⁵.

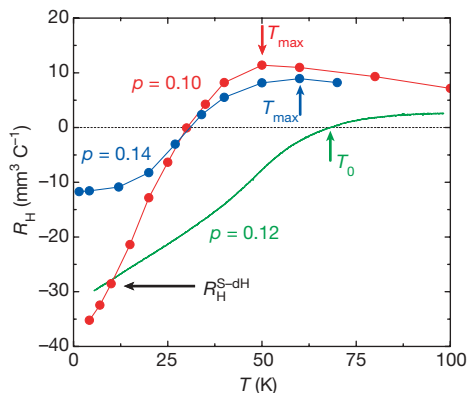


Figure 3 | Normal-state Hall coefficient versus temperature. Hall coefficient R_H versus T for Y123-II, Y123-VIII and Y124 (data multiplied by ten), at $B = 55$, 45 and 55 T, respectively. T_0 is the temperature at which R_H changes sign, equal to 30, 70 and 30 ± 2 K, respectively. T_{max} is the temperature at which R_H is maximum, equal to 50, 105 and 60 ± 5 K, respectively. The black arrow indicates the value of the Hall coefficient expected for a single electron Fermi pocket of the size imposed by Shubnikov–de Haas oscillations of frequency F , namely $R_H^{\text{S-dH}} = -V_{\text{cell}}/en_{\text{S-dH}}$, where $n_{\text{S-dH}} = F/\Phi_0 = 0.038$ electrons per unit cell. (The data for Y124 is multiplied by a factor of ten to put it on a scale comparable to Y123-II and Y123-VIII. The order-of-magnitude reduction of the measured Hall voltage comes in large part from the short-circuiting effect of the CuO chains along the b axis, which in this stoichiometric material, unlike in Y123-II and Y123-VIII, remain highly conductive down to low temperature; see text.)

Supplementary Information. The 4.2-K isotherm of Y124 illustrates nicely the basic components of R_H : the flat negative part at high field (above B_n) is the normal-state value, whereas the positive overshoot just above B_s is due to a vortex flux-flow contribution.

The most natural explanation for the negative R_H is the presence of an electron pocket in the Fermi surface. (In principle, it could also come from a hole pocket with portions of negative curvature¹⁶.) In a scenario in which the Fermi surface contains both electron and hole pockets, the sign of R_H depends on the relative magnitude of the respective densities n_e and n_h , and mobilities¹⁷ μ_e and μ_h . ($\mu \equiv e\tau/m^*$, where e is the electron charge, τ^{-1} is the scattering rate and m^* is the effective mass.) Given that these materials are hole-doped, we expect $n_h > n_e$. The fact that $R_H < 0$ at low T therefore implies that $\mu_e > \mu_h$ at low T . Given strong inelastic scattering, this inequality can then easily invert at high T , offering a straightforward mechanism for the sign change in R_H . This happens in simple metals like Al and In (ref. 17) and is typical of compensated metals (with $n_e = n_h$). In high-purity samples of NbSe₂, a quasi-two-dimensional metal that undergoes a charge density-wave transition at $T_{\text{CDW}} \approx 30$ K, $R_H(T)$ drops from its positive and flat behaviour above T_{CDW} eventually to become negative below $T_0 \approx 25$ K (ref. 18), as reproduced in Supplementary Fig. 5. In impure samples, however, $R_H(T)$ remains positive at all T values (ref. 18; Supplementary Fig. 5), showing that the electron/hole balance can depend sensitively on impurity/disorder scattering.

A scenario of electron and hole pockets for YBCO resolves a puzzle in relation to the Shubnikov–de Haas oscillations observed in Y123-II (ref. 2). The puzzle is the apparent violation of the Luttinger sum rule, which states that the total carrier density n must be equal to the total area of the two-dimensional Fermi surface. From the oscillation frequency $F = 530$ T, one gets a carrier density $n_{\text{S-dH}} = 0.038$ carriers per planar Cu atom per pocket via $F = n_{\text{S-dH}}\Phi_0$, where $\Phi_0 = 2.07 \times 10^{-15} \text{ T m}^2$ is the flux quantum. Assuming that the pocket is a hole pocket (of arbitrary curvature) and there is nothing else in the Fermi surface, and assuming also that n must be equal to the density of doped holes ($n = p = 0.10$), the Luttinger sum rule is clearly violated, whether the relevant Brillouin zone includes one or two (or any number) of these pockets¹⁹ (whether $n = n_{\text{S-dH}} = 0.038$ or $n = 2n_{\text{S-dH}} = 0.076$). If, on the other hand, the Fermi surface contains other sheets (not seen in the Shubnikov–de Haas oscillations) besides the observed pockets, then the sum rule can easily be satisfied.

The fact that R_H is negative at low T implies that the Shubnikov–de Haas frequency that was seen in Y123-II (ref. 2) must come from the high-mobility electron pocket, because the amplitude of Shubnikov–de Haas oscillations depends exponentially on mobility, as $\exp(-\pi/\mu B)$. The hole-like portions of the Fermi surface are either open or have a lower mobility at $T \rightarrow 0$. The largest value of R_H that a single electron pocket of density $n_{\text{S-dH}} = 0.038$ electrons per unit cell can produce is $R_H^{\text{S-dH}} = -V_{\text{cell}}/en_{\text{S-dH}} = -29 \text{ mm}^3 \text{C}^{-1}$. Within the uncertainty in the geometric factor, this is the magnitude of R_H

measured in Y123-II and Y123-VIII at low temperature (see Fig. 3). (A similar estimate cannot be made for Y124, because in this particular case, unlike in Y123-II and Y123-VIII with their imperfect CuO chains that localize charge carriers², the chains remain metallic down to low temperature, and thus partially short-circuit the Hall voltage, causing a reduction in Hall resistance by as much as a factor of ten or so (see ref. 15).) The picture that emerges is in sharp contrast with the single Fermi 'arc' at $(\pi/2, \pi/2)$ seen in angle-resolved photo-emission spectroscopy (ARPES) studies on other copper oxides in zero field (see discussion in refs 1 and 2). A possible explanation for the discrepancy is that ARPES detects only the hole pocket (one side of it) and the Shubnikov-de Haas oscillations only the electron pocket (thanks to its high mobility in the elastic scattering regime). This would suggest a Fermi surface similar to that proposed for electron-doped copper oxides (see below), with a hole pocket at $(\pi/2, \pi/2)$ and an electron pocket at $(\pi, 0)$.

One might ask why a negative R_H has not been seen in $\text{Bi}_2\text{Sr}_{2-x}\text{La}_x\text{CuO}_{6+\delta}$ (BSLCO) and $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ (LSCO), the other hole-doped copper oxides to have been measured up to 50 T (refs 20 and 21). Indeed, although R_H can drop by nearly a factor of three between 150 and 1.5 K in BSLCO, in a manner not unlike what is seen here in YBCO, it never becomes negative. A possible explanation is that the negative R_H in the YBCO copper oxides is associated with the presence of CuO chains. In Y123-II, however, electrical anisotropy in the a - b plane is unity below 100 K (ref. 2), implying that the chain subsystem is non-conducting at low temperatures. In Y124, the double-chain unit remains metallic down to low T and will therefore have an associated Hall coefficient. However, in isostructural, non-superconducting Pr124, in which only the chains are conducting, R_H is found to be positive at low T (ref. 22). A more likely alternative is that the much stronger disorder scattering characteristic of BSLCO and LSCO compared to Y123 or Y124 suppresses μ_e more severely than μ_h . In such a case, the electron pocket, although present, manifests itself only in the temperature dependence of R_H , not its sign. As mentioned above, this is what happens in NbSe_2 : adding impurities eliminates the sign change in R_H (see ref. 18 and Supplementary Fig. 5).

Because the band structures of Y123 (see ref. 2 and references therein) and Y124 (see ref. 4 and references therein) calculated within the local density approximation do not support electron pockets, the fundamental question is: how do electron pockets come to exist? The combination of a small Fermi surface volume from Shubnikov-de Haas oscillations and a negative R_H pointing to electron pockets, in both Y123 and Y124, argues strongly for a reconstruction of the local density approximation Fermi surface. The standard mechanism for such reconstruction is the onset of a density-wave instability, as encountered in numerous materials. Examples in quasi-two-dimensional materials include NbSe_2 , already mentioned, and the transition-metal oxide $\text{Ca}_3\text{Ru}_2\text{O}_7$ (ref. 23).

In copper oxides, at least three mechanisms can be invoked for a reconstruction of the Fermi surface that would result in electron and hole pockets. The first scenario is an antiferromagnetic phase with a (π, π) ordering wavevector, which causes the large hole-like Fermi surface to reconstruct into a small hole pocket at $(\pi/2, \pi/2)$ and a small electron pocket at $(\pi, 0)$. This model was used to explain the sign change in the low-temperature R_H measured in the electron-doped copper oxide $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_{4-\delta}$ (PCCO) (ref. 7), on crossing a critical concentration x_c close to where long-range antiferromagnetic order ends. In PCCO, R_H is positive at all temperatures at high doping ($x > 0.19$) and negative at low doping ($x < 0.15$), but it changes sign at intermediate dopings, from positive below $T_0 = 30$ –40 K to negative above⁶. The latter behaviour is similar (but opposite in sign) to that of YBCO in the range $0.10 < p < 0.14$, suggesting that the transport properties of both materials should be interpreted in terms of electrons and holes with different, T -dependent mobilities. In the case of hole-doped copper oxides, however, it is less likely that antiferromagnetic order is the relevant

mechanism, because long-range antiferromagnetic order is thought to be confined to lower doping ($p < 0.05$). Nevertheless, the possibility should be investigated, particularly as a large magnetic field may push the phase transition to higher doping¹⁹.

The second scenario is a theoretical phase with d -density-wave order, which would also cause a (π, π) folding of the Fermi surface and thus could produce electron pockets near $(\pi, 0)$ (ref. 24). The third scenario is a density-wave phase akin to that encountered experimentally in LBCO and Nd-doped LSCO. In these systems, a Fermi surface reconstruction, signalled by a precipitous drop in R_H (Fig. 1), coincides with the density-wave transition, observed directly via neutron diffraction. Reference 25 has recently shown that within mean-field theory a '1/8-stripe' spin/charge density-wave order of this kind does reconstruct the Fermi surface of a generic hole-doped copper oxide in a way that tends to produce an electron pocket.

Although the similarities between the La_2CuO_4 and YBCO families are highly suggestive, there are some important differences. First, whereas there is unambiguous direct evidence for static spin/charge density-wave order in LBCO and Nd-LSCO from neutron diffraction^{8,9}, there is no such evidence so far in YBCO. This could be because the putative density-wave phase in YBCO involves fluctuating rather than static order²⁶, or short-range rather than long-range order²⁷. Second, the anomalies in the transport properties associated with density-wave order are sharp in LBCO (ref. 11) and Nd-doped LSCO (ref. 10), whereas in YBCO the temperature dependence of R_H is smooth. Note, however, that anomalies in the former materials seem to be sharp only when the low-temperature structural transition coincides with the density-wave transition. When structural and density-wave transitions do not coincide, as in Eu-doped LSCO (ref. 28), the anomalies also seem to be smooth.

In summary, the normal state of underdoped YBCO is characterized by a negative Hall coefficient, revealing the presence of an electron pocket in the Fermi surface, the mobility of which in these clean copper oxides is high enough to outweigh the contribution from other, hole-like parts of the Fermi surface. This implies that the Shubnikov-de Haas oscillations observed recently in the same materials must come from those electron pockets. It also suggests that the generally positive Hall coefficient seen in other hole-doped copper oxides results from electron mobilities that are too low because of stronger disorder scattering. Because electron pockets are not supported by the band structure of YBCO, we conclude that they must come from a reconstruction of the Fermi surface, which occurs at a critical doping above $p = 0.14$. In the absence of any direct evidence so far for long-range density-wave order in YBCO, our findings call for theoretical investigations of more unconventional scenarios.

METHODS SUMMARY

Samples. The Y123 samples are fully detwinned crystals of $\text{YBa}_2\text{Cu}_3\text{O}_y$, grown in non-reactive BaZrO_3 crucibles from high-purity starting materials (see ref. 2 and references therein). For Y123-II (or Y123-VIII), the oxygen content was set at $y = 6.51$ (or 6.67) and the dopant oxygen atoms were made to order into an ortho-II (or ortho-VIII) superstructure, yielding a superconducting transition temperature $T_c = 57.5$ K (or 66.0 K). The samples are uncut, unpolished thin platelets, the transport properties of which are measured via gold evaporated contacts (resistance $< 1 \Omega$), in a six-contact geometry. Typical sample dimensions are 20 – 50×500 – 800×500 – $1,000 \mu\text{m}^3$ (thickness \times width \times length). The $\text{YBa}_2\text{Cu}_4\text{O}_8$ crystals were grown by a flux method in Y_2O_3 crucibles and an Ar/O_2 mixture at 2,000 bar, with a partial oxygen pressure of 400 bar (ref. 29).

Estimates of hole doping. The hole doping p in Y123 is determined from a relationship between T_c and the c axis lattice constant³⁰. For our Y123-II (or Y123-VIII) samples, the measured T_c implies $p = 0.099$ (or 0.120). We assume the doping in Y124 is the same as in a sample of Y123 with the same T_c of 80 K, namely $p = 0.137 \approx 0.14$.

Resistance measurements. Longitudinal (R_{xx}) and transverse (R_{xy}) resistances are obtained from the voltage drop measured diagonally on either side of the sample width, for a field parallel (up) and anti-parallel (down) to the c axis: $R_{xx} \equiv (V_{\text{up}} + V_{\text{down}})/2I_x$ and $R_{xy} \equiv (V_{\text{up}} - V_{\text{down}})/2I_x$. Measurements on Y123-II and Y124 were performed at the LNCMP in Toulouse, in a pulsed resistive

magnet up to 61 T. Measurements on Y123-II and Y123-VIII were performed at the NHMFL in Tallahassee, in a steady hybrid magnet up to 45 T.

Received 31 July; accepted 25 September 2007.

1. Julian, S. R. & Norman, M. R. Local pairs and small surfaces. *Nature* **447**, 537–539 (2007).
2. Doiron-Leyraud, N. *et al.* Quantum oscillations and the Fermi surface in an underdoped high- T_c superconductor. *Nature* **447**, 565–568 (2007).
3. Yelland, E. A. *et al.* Quantum oscillations in the underdoped cuprate $\text{YBa}_2\text{Cu}_4\text{O}_8$. Preprint at (<http://arXiv.org/abs/0707.0057>) (2007).
4. Bangura, A. F. *et al.* Shubnikov-de Haas oscillations in $\text{YBa}_2\text{Cu}_4\text{O}_8$. Preprint at (<http://arXiv.org/abs/0707.4461>) (2007).
5. Hussey, N. E. *et al.* Observation of a coherent three-dimensional Fermi surface in a high-transition temperature superconductor. *Nature* **425**, 814–817 (2003).
6. Li, P., Balakirev, F. F. & Greene, R. L. High-field Hall resistivity and magneto-resistance in electron-doped $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_{4-\delta}$. *Phys. Rev. Lett.* **99**, 047003 (2007).
7. Lin, J. & Millis, A. J. Theory of low-temperature Hall effect in electron-doped cuprates. *Phys. Rev. B* **72**, 214506 (2005).
8. Tranquada, J. M. *et al.* Evidence for stripe correlations of spins and holes in copper oxide superconductors. *Nature* **375**, 561–563 (1995).
9. Ichikawa, N. *et al.* Local magnetic order vs superconductivity in a layered cuprate. *Phys. Rev. Lett.* **85**, 1738–1741 (2000).
10. Noda, T., Eisaki, H. & Uchida, S. Evidence for one-dimensional charge transport in $\text{La}_{2-y}\text{xNd}_y\text{Sr}_x\text{CuO}_4$. *Science* **286**, 265–268 (1999).
11. Adachi, T., Noji, T. & Koike, Y. Crystal growth, transport properties, and crystal structure of the single-crystal $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$ ($x=0.11$). *Phys. Rev. B* **64**, 144524 (2001).
12. Harris, J. M. *et al.* Hall angle evidence for the superclean regime in 60-K $\text{YBa}_2\text{Cu}_3\text{O}_{6+y}$. *Phys. Rev. Lett.* **73**, 1711–1714 (1994).
13. Ito, T., Takenaka, K. & Uchida, S. Systematic deviation from T -linear behavior in the in-plane resistivity of $\text{YBa}_2\text{Cu}_3\text{O}_{7-y}$: evidence for dominant spin scattering. *Phys. Rev. Lett.* **70**, 3995–3998 (1993).
14. Wang, Y. & Ong, N. P. Particle-hole symmetry in the antiferromagnetic state of the cuprates. *Proc. Natl Acad. Sci. USA* **98**, 11091–11096 (2001).
15. Segawa, K. & Ando, Y. Intrinsic Hall response of the CuO_2 planes in a chain-plane composite system of $\text{YBa}_2\text{Cu}_3\text{O}_y$. *Phys. Rev. B* **69**, 104521 (2004).
16. Ong, N. P. Geometric interpretation of the weak-field Hall conductivity in two-dimensional metals with arbitrary Fermi surface. *Phys. Rev. B* **43**, 193–201 (1991).
17. Ashcroft, N. W. The reversal of Hall fields in aluminium and indium. *Phys. Kondens. Mater.* **9**, 45–53 (1969).
18. Huntley, D. J. & Frindt, R. F. Transport properties of NbSe_2 . *Can. J. Phys.* **52**, 861–867 (1974).
19. Chen, W.-Q., Yang, K.-Y., Rice, T. M. & Zhang, F.-C. Quantum oscillations in magnetic-field-induced antiferromagnetic phase of underdoped cuprates: application to ortho-II $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$. Preprint at (<http://arXiv.org/abs/0706.3556>) (2007).
20. Balakirev, F. F. *et al.* Signature of optimal doping in Hall-effect measurements on a high-temperature superconductor. *Nature* **424**, 912–915 (2003).
21. Balakirev, F. F. *et al.* Magneto-transport in LSCO high- T_c superconducting thin films. *N. J. Phys.* **8**, 194 (2006).
22. Horii, S. *et al.* On the dimensionality of the Cu-O double-chain site of $\text{PrBa}_2\text{Cu}_4\text{O}_8$. *Phys. Rev. B* **66**, 054530 (2002).
23. Baumberger, F. *et al.* Nested Fermi surface and electronic instability in $\text{Ca}_3\text{Ru}_2\text{O}_7$. *Phys. Rev. Lett.* **96**, 107601 (2006).
24. Chakravarty, S. *et al.* Sharp signature of a $d_{x^2-y^2}$ quantum critical point in the Hall coefficient of cuprate superconductors. *Phys. Rev. Lett.* **89**, 277003 (2002).
25. Millis, A. J. & Norman, M. R. Antiphase stripe order as the origin of electron pockets observed in 1/8-hole-doped cuprates. Preprint at (<http://arXiv.org/abs/0709.0106>) (2007).
26. Kivelson, S. A. *et al.* How to detect fluctuating stripes in the high-temperature superconductors. *Rev. Mod. Phys.* **75**, 1201–1241 (2003).
27. Kohsaka, Y. *et al.* An intrinsic bond-centered electronic glass with unidirectional domains in underdoped cuprates. *Science* **315**, 1380–1385 (2007).
28. Hücker, M. *et al.* Consequences of stripe order for the transport properties of rare earth doped $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. *J. Phys. Chem. Solids* **59**, 1821–1824 (1998).
29. Adachi, S. *et al.* Preparation of $\text{YBa}_2\text{Cu}_4\text{O}_8$ single crystals in Y_2O_3 crucible using O_2 -HIP apparatus. *Physica C* **301**, 123–128 (1998).
30. Liang, R., Bonn, D. A. & Hardy, W. N. Evaluation of CuO_2 plane hole doping in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ single crystals. *Phys. Rev. B* **73**, 180505 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. W. Ashcroft, K. Behnia, L. Brisson, S. Chakravarty, J. C. Davis, R. L. Greene, S. A. Kivelson, G. G. Lonzarich, M. R. Norman, A. J. Schofield, A.-M. S. Tremblay and D. Vignolle for discussions, and J. Corbin and M. Nardone for their help with the experiments. We acknowledge support from the Canadian Institute for Advanced Research, the LNCMP and the NHMFL, and funding from the NSERC, the FQRNT, the EPSRC and a Canada Research Chair. Part of this work was supported by the French ANR IceNET and EuroMagNET. The NHMFL is supported by an NSF grant and the State of Florida.

Author Contributions D.L. and N.D.-L. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.P. (proust@lncmp.org) or L.T. (louis.taillefer@physique.usherbrooke.ca).

Dissolved organic carbon trends resulting from changes in atmospheric deposition chemistry

Donald T. Monteith^{1*}, John L. Stoddard^{2*}, Christopher D. Evans³, Heleen A. de Wit⁴, Martin Forsius⁵, Tore Høgåsen⁴, Anders Wilander⁶, Brit Lisa Skjelkvåle⁴, Dean S. Jeffries⁷, Jussi Vuorenmaa⁵, Bill Keller⁸, Jiri Kopáček⁹ & Josef Vesely^{10‡}

Several hypotheses have been proposed to explain recent, widespread increases in concentrations of dissolved organic carbon (DOC) in the surface waters of glaciated landscapes across eastern North America and northern and central Europe^{1–3}. Some invoke anthropogenic forcing through mechanisms related to climate change^{3–5}, nitrogen deposition⁶ or changes in land use⁷, and by implication suggest that current concentrations and fluxes are without precedent. All of these hypotheses imply that DOC levels will continue to rise, with unpredictable consequences for the global carbon cycle. Alternatively, it has been proposed that DOC concentrations are returning toward pre-industrial levels as a result of a gradual decline in the sulphate content of atmospheric deposition^{8–10}. Here we show, through the assessment of time series data from 522 remote lakes and streams in North America and northern Europe, that rising trends in DOC between 1990 and 2004 can be concisely explained by a simple model based solely on changes in deposition chemistry and catchment acid-sensitivity. We demonstrate that DOC concentrations have increased in proportion to the rates at which atmospherically deposited anthropogenic sulphur and sea salt have declined. We conclude that acid deposition to these ecosystems has been partially buffered by changes in organic acidity and that the rise in DOC is integral to recovery from acidification. Over recent decades, deposition-driven increases in organic matter solubility may have increased the export of DOC to the oceans, a potentially important component of regional carbon balances¹¹. The increase in DOC concentrations in these regions appears unrelated to other climatic factors.

A number of national monitoring programmes, designed to assess the effects of acid emission reductions on acid-sensitive lakes and streams, have reported increases in DOC concentrations over the past two decades^{2,4,9}. Many monitoring networks were initiated in the late 1980s, and our analysis focuses on the period 1990–2004 to maximize geographical coverage. Over this period, sulphur deposition fell sharply in most regions². Simultaneously, sea salt deposition to some coastal regions of northwestern Europe declined, following prolonged exceptionally stormy conditions in the early 1990s (ref. 12).

The geographical extent of recorded DOC increases suggests a common dominant driver, and highlights the need for a large-scale inter-regional analysis of time series data¹. We examined the spatial distribution of DOC trends in data collected in glaciated regions of six North European and North American countries, and assessed the mechanisms capable of explaining observed patterns.

Our trend results are consistent with previous reports of DOC increases in surface waters on glaciated, acid-sensitive terrain in Europe and North America (Fig. 1; also see Supplementary Information). Upward slopes ($n = 363$) outnumbered downward slopes ($n = 139$), and 88% of significant trends ($P < 0.05$) were positive. Upward slopes were most frequently significant in the Southern Nordic region (below 62° latitude), in the UK, and in the northeastern USA. Between 1990 and 2004, sulphur and/or chloride deposition declined in these regions as a result of reductions in sulphur emissions (all areas) and in the severity of coastal storms, respectively. The majority of DOC records in the northern Nordic region, and in Ontario/Quebec, also showed positive slopes, although fewer of these were significant ($P < 0.05$). Atlantic Canada was the only region with little evidence of increasing DOC.

We found that relative change in DOC (that is, DOC change per year as a percentage of the site median, or % Δ DOC) was strongly and inversely related to change in both sulphate concentration (ΔSO_4^{2-}) and chloride concentration (ΔCl^-) ($n = 522$; Fig. 2). Surface water SO_4^{2-} and Cl^- concentrations in these northern ecosystems are generally highly correlated with concentrations of these ions in deposition, and can serve as surrogates for site-specific measurements of SO_4^{2-} and Cl^- deposition¹³ (see Methods Summary). While falling SO_4^{2-} dominated most records, large DOC increases occurred in some coastal locations where ΔSO_4^{2-} was small, provided that ΔCl^- was strongly downward. Increased sea salt deposition appeared to be critical in explaining the unusual DOC declines in some parts of Atlantic Canada (Fig. 1).

To explore potential factors affecting the % Δ DOC, we performed a forward-selection stepwise regression that included variables representing both trend results (annual slopes) and median values of key chemical and regional temperature parameters (see Methods Summary). ΔSO_4^{2-} and ΔCl^- were selected, along with median non-marine $\text{Ca}^{2+} + \text{Mg}^{2+}$ concentration (CaMg^* , representing catchment sensitivity to acidification), and an interaction term ($\Delta\text{Cl}^- \times \text{CaMg}^*$) (Table 1). All four variables were highly significant, and explained 39.3% of the variance in % Δ DOC. Given the uncertainties surrounding trend estimation, and the influence of inevitable local and sub-regional factors on DOC trends at individual sites, the explanatory power of this single multi-regional model is considerable.

Residuals from our model support the idea that recent widespread DOC changes are related to changes in the atmospheric deposition of SO_4^{2-} and Cl^- across large areas (Fig. 3). The UK was the only region where the median of residuals was significantly different from zero, perhaps because the true trend in SO_4^{2-} deposition for UK sites was

¹Environmental Change Research Centre, UCL, London, WC1E 6BT, UK. ²US EPA, Corvallis, Oregon 97333, USA. ³Centre for Ecology and Hydrology, Bangor, LL57 2UW, UK.

⁴Norwegian Institute for Water Research, N-0349 Oslo, Norway. ⁵Finnish Environment Institute, PO Box 140, FI-00251 Helsinki, Finland. ⁶Department of Environment Assessment SLU, SE-75007 Uppsala, Sweden. ⁷Environment Canada, Burlington, Ontario, L7R4A6, Canada. ⁸Ontario Ministry of the Environment, Laurentian University, Sudbury, Ontario, P3E 2C6, Canada. ⁹Biology Centre, Institute of Hydrobiology, 37005 České Budějovice, Czech Republic. ¹⁰Czech Geological Survey, 152 00 Prague, Czech Republic.

*These authors contributed equally to this work.

‡Deceased.

underestimated (owing to short-term adsorption of SO_4^{2-} during heightened sea salt deposition in the early 1990s in this region, the most maritime in the data set¹²). The model can account entirely for the average slight negative DOC trend in Atlantic Canada, the small average increases in the northeastern USA, Ontario/Quebec and northern Nordic regions and the large average increase in the southern Nordic region. The approximately even distribution of residuals around the median for each region suggests that much of the residual variance is of a local or random, rather than systematic, nature (Fig. 3).

A number of other factors are known to influence short-term variation in DOC concentration in surface waters, and some of those that have undergone recent change have been proposed to explain long-term DOC trends^{3–7,14,15}. None of these factors, however, exhibit spatial patterns that are consistent with the patterns of DOC change that we observe (Fig. 1 and Supplementary Information). For example, recent data from rivers in the UK¹⁶, the eastern USA¹⁷, Norway¹⁸, Canada¹⁹ and Finland²⁰ provide no evidence for trends in seasonal or annual hydrology that are spatially consistent with the distribution of DOC trends. We could not use runoff chemistry to assess the influence of changing nitrogen (N) deposition on DOC trends, because atmospheric N is retained strongly in most catchments and is thus non-conservative. Although N is certainly accumulating in many ecosystems, both the amounts of N deposition, and temporal trends in N deposition and runoff, are highly variable across our sites²¹. In our data set, only 28 sites showed significant ($P < 0.05$) increases in nitrate concentration (indicative of possible nitrogen saturation), 100 showed declines and 394 showed no trend. We found no indication of trends in seasonal or annual air temperatures across northern Europe during the period 1990–2004, with the exception of Scotland (autumn only) and northern and eastern Finland (summer only). Upward temperature trends were more widespread in eastern North America, but occurred in regions where there was no dominant DOC tendency. Although CO_2 concentrations are increasing globally, they cannot simultaneously explain both the upward and downward trends in DOC we observe in different regions. Finally, there have been no consistent land-use changes across all regions.

Changes in the chemistry of atmospheric deposition, therefore, provided the only regionally consistent explanation for widespread upward trends in surface water DOC concentrations. A comparison of the relative effects of non-marine (that is, anthropogenic) SO_4^{2-} and the combined effects of Cl^- and marine SO_4^{2-} , using the coefficients from our model, showed that, on average, the reduction in anthropogenic SO_4^{2-} was responsible for >85% of the total anion effect on ΔDOC in all regions other than the UK (where anthropogenic and marine effects were similar) and Atlantic Canada (where marine effects dominated) (see Supplementary Information). However, the selection of ΔCl^- in the model, with a coefficient of similar magnitude to that for ΔSO_4^{2-} , is suggestive of common soil processes.

Atmospheric deposition can affect soil organic matter (SOM) solubility through at least two mechanisms—by changing either the acidity of soils or the ionic strength of soil solutions, or both. Acid deposition and sea salt can exert similar short-term effects on both processes. Krug and Frink²² originally proposed that increasing the pH of base-poor soils should increase SOM solubility. Although their premise—that surface water acidification would be minimized by the resulting substitution of mineral acids for organic acids—has been widely disproved, SOM solubility has since been linked to soil pH in laboratory^{23,24} and field^{23,25} studies; soil equilibrium models²⁶ now include the pH dependence of SOM solubility. Soil pH also determines the solubility of aluminium which, when released during the process of soil acidification, can bind with organic molecules, neutralising high-affinity binding-sites and thereby triggering SOM coagulation²⁷.

Declining acid deposition and, in a minority of sites, a decline in the frequency and intensity of sea salt episodes may affect DOC concentrations both directly, by increasing soil pH, and indirectly, by reducing aluminium mobilisation. The selection of CaMg^* , an indicator of acid sensitivity in our multiple regression model (Table 1), provides some support for this soil pH/DOC solubility mechanism.

A second mechanism involves the effect of ionic strength on the coagulation of DOC. Decreasing inputs of acidic deposition or sea salt should lead to lower concentrations of a suite of multivalent ions found in soil solution—including SO_4^{2-} , Ca^{2+} , Mg^{2+} and aluminium—that

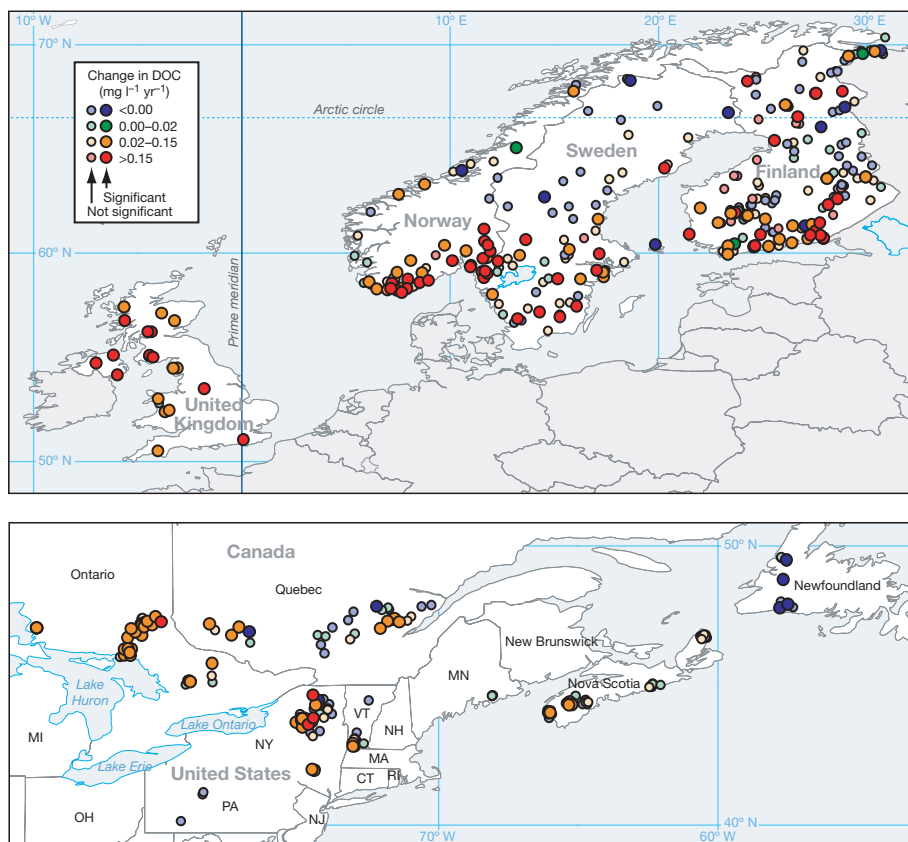


Figure 1 | Trends in dissolved organic carbon ($\text{mg l}^{-1} \text{yr}^{-1}$). Data are shown for monitoring sites on acid-sensitive terrain in Europe (upper panel) and North America (lower panel) for the period 1990–2004.

exert a large influence on ionic strength. Previous studies have shown that increasing the ionic strength of soil solutions reduces the rate of DOC flux^{26,28}; conversely, a reduction in ionic strength (for example, through decreasing atmospheric deposition) should lead to a DOC increase. The selection of CaMg* in our model may also be relevant here, because the ionic strength of the soil solution in catchments where CaMg* is low should be more sensitive to change in the deposition of these ions.

Evidence for our conclusions thus derives from three lines of support: (1) a plausible mechanism; (2) experimental studies showing

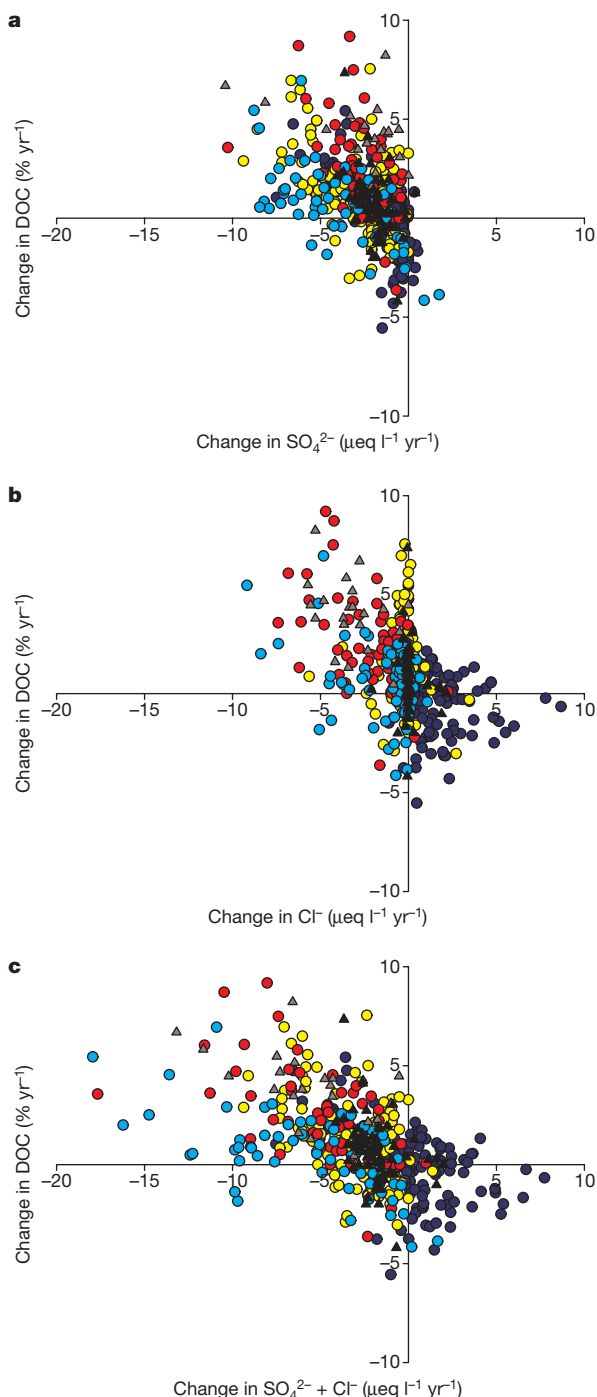


Figure 2 | Relationship between %ΔDOC, ΔSO_4^{2-} and ΔCl^- and the equivalent sum of ΔSO_4^{2-} and ΔCl^- , used as surrogates for changes in atmospheric deposition. a, ΔSO_4^{2-} ; b, ΔCl^- ; c, ΔSO_4^{2-} and ΔCl^- . Symbols represent data from Canada (dark blue circles), Finland (yellow circles), Norway (red circles), Sweden (light blue circles), the UK (grey triangles) and the USA (black triangles).

Table 1 | Results of stepwise multiple regression for %ΔDOC

Variable	Estimate	P	Cumulative R^2
Intercept	0.250	0.0004	NA
Change in SO_4^{2-} ($\mu\text{eq l}^{-1} \text{yr}^{-1}$)	-0.557	<0.0001	0.237
Median CaMg* ($\mu\text{eq l}^{-1}$)	-0.005	<0.0001	0.316
Change in Cl^- ($\mu\text{eq l}^{-1} \text{yr}^{-1}$)	-0.504	<0.0001	0.360
Median CaMg* ($\mu\text{eq l}^{-1}$) \times change in Cl^- ($\mu\text{eq l}^{-1} \text{yr}^{-1}$)	0.0033	<0.0001	0.393

See Methods Summary. Estimates for intercept and slope of individual relationships are for the final three variable plus one interaction model. NA, not applicable. μeq , microequivalents. *Sea-salt-corrected divalent cation concentration.

that our proposed mechanism is capable of causing the observed DOC response; and (3) temporal and spatial coherence of both the 'cause' (changing atmospheric deposition) and the 'effect' (changing DOC levels). Our results provide the first clear, large-scale demonstration that DOC concentrations are sensitive to changes in acid deposition, and, therefore, that declining surface water acidity (mediated by decreasing acid deposition) is being partially offset by increasing organic acidity. Furthermore, they suggest that DOC fluxes from these regions to the oceans may be returning to levels more typical of pre-industrial times.

In maritime northwestern Europe, both SO_4^{2-} and Cl^- have fallen over the 1990–2004 period, but for unrelated reasons. Sea salt deposition varies due to meteorological factors that affect Atlantic storminess, such as the North Atlantic Oscillation¹², producing roughly decadal oscillations. Trends in SO_4^{2-} deposition, on the other hand, have operated over much longer timescales and are similar to those in other regions covered by this study. Rates of sulphur deposition began rising in the mid- to late 1800s, peaked in the 1970s, and have subsequently fallen. At timescales greater than decades, long-term trends in DOC will be controlled by changes in sulphate deposition, even in regions subject to large sea salt inputs.

Although acid deposition has affected large parts of Europe and North America, it has not been a global phenomenon. These findings emphasize, therefore, that evidence of rising DOC in these (well-studied) areas should not be misconstrued as evidence of rising DOC

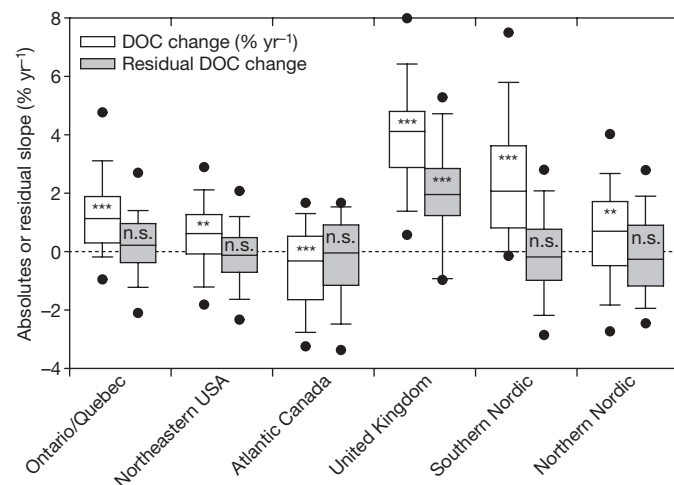


Figure 3 | Distributions of %ΔDOC and residuals from multiple regression, by region. Boxplots are of regional distributions of $\Delta\% \text{DOC}$ per year (unshaded boxes) and residuals calculated from a multiple regression predicting $\Delta\% \text{DOC}$ from ΔSO_4^{2-} , ΔCl^- , median CaMg* concentration and the interaction term $\Delta\text{Cl}^- \times \text{CaMg}^*$ (grey boxes). Symbols within boxes indicate significance of the one-sample median test of H_0 : median = 0 (*** P < 0.001; ** P < 0.01; not significant P > 0.05). Boundaries of boxes indicate interquartile range (25th to 75th percentiles) and median value (midline); whiskers indicate 10th and 90th percentiles and symbols indicate 5th and 95th percentiles. Residual values centre around zero for all regions, with the possible exception of the UK, indicating that the regression model built on all data reliably explains the tendency for $\Delta\% \text{DOC}$ concentrations to deviate from zero in the separate regions.

in unmonitored waters globally. They also suggest that threats of widespread destabilization of terrestrial carbon reserves by gradual rises in air temperature or CO₂ concentration^{3–5} may have been overstated. There is, however, a clear need for closer integration of research into pollutant and climatic impacts on terrestrial carbon dynamics. An improved understanding of the roles of both, and their interactions, is required to predict future surface water quality accurately, and to determine crucially important linkages between climate and the carbon cycle.

METHODS SUMMARY

Water chemistry data were collated from several regional and national monitoring initiatives for the period 1990–2004. Records represent mostly headwater lakes and lower-order streams that are largely free of local disturbance, other than potential effects of forestry and rough grazing practices. Data were analysed for trends by the Mann–Kendall test²⁹, with slopes estimated by the Sen estimator³⁰.

Site-specific deposition estimates were not available for most of the surface water monitoring sites, so we used changes in measured surface water concentrations of SO₄^{2–} (ΔSO₄^{2–}) and Cl[–] (ΔCl[–]), as deposition surrogates. Inputs and outputs of SO₄^{2–} rarely balance precisely, but in younger (that is, glaciated) northern ecosystems the two are generally highly correlated¹³.

Regional trends in seasonal (that is, mean December–February, March–May, June–August and September–November) and annual air temperature were determined from CRUTEM2 simulated air temperature anomaly data from the Climate Research Unit, University of East Anglia (<http://www.cru.uea.ac.uk/cru/data/tem2/>). These data are compiled in a 5° × 5° global grid, so several monitoring sites may be associated with a single, grid-based estimate of temperature change over the 1990–2004 time period relative to the 1961–1990 average.

To determine factors influencing %ADOC at the regional scale, we conducted a stepwise multiple regression using, for each site, both trend results (annual slopes) and median values of key chemical variables (SO₄^{2–}, non-marine SO₄^{2–}, Cl[–], NO₃[–], H⁺, acid-neutralizing capacity, Gran alkalinity, base cations, CaMg* and aluminium), together with slopes for annual and seasonal temperature anomalies for the representative 5° grid square for each site. In an initial run, ΔSO₄^{2–}, ΔCl[–] and CaMg* were the only variables selected. Because sites with declining ΔSO₄^{2–} and ΔCl[–] dominated the data set, the negative coefficient for CaMg* (Table 1) suggested that the DOC response to declining anion concentrations was greatest in the most acid-sensitive systems. We therefore introduced an additional interaction term (ΔCl[–] × CaMg*) to account for the potentially dampening effect of larger base cation concentrations on negative DOC trends at a minority of sites at which ΔCl[–] had increased.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 December 2006; accepted 24 September 2007.

1. Roulet, N. & Moore, T. R. Browning the waters. *Nature* **444**, 283–284 (2006).
2. Skjelkvåle, B. L. et al. Regional scale evidence for improvements in surface water chemistry 1990–2001. *Environ. Pollut.* **137**, 165–176 (2005).
3. Worrall, F., Burt, T. & Shedden, R. Long term records of riverine dissolved organic matter. *Biogeochemistry* **64**, 165–178 (2003).
4. Freeman, C., Evans, C. D., Monteith, D. T., Reynolds, B. & Fenner, N. Export of organic carbon from peat soils. *Nature* **412**, 785 (2001).
5. Freeman, C. et al. Export of dissolved organic carbon from peatlands under elevated carbon dioxide levels. *Nature* **430**, 195–198 (2004).
6. Findlay, S. E. G. Increased carbon transport in the Hudson River: unexpected consequence of nitrogen deposition? *Frontiers Ecol. Environ.* **3**, 133–137 (2005).
7. Garnett, M. H., Ineson, P. & Stevenson, A. C. Effects of burning and grazing on carbon sequestration in a Pennine blanket bog. *Holocene* **10**, 729–736 (2000).
8. Evans, C. D., Chapman, P. J., Clark, J. M., Monteith, D. T. & Cresser, M. S. Alternative explanations for rising dissolved organic carbon export from organic soils. *Glob. Change Biol.* **12**, 1–10 (2006).
9. Stoddard, J. L. et al. *Response of Surface Water Chemistry to the Clean Air Act Amendments of 1990*. Report EPA/620/R-03/001 (US Environmental Protection Agency, Washington DC, 2003).
10. Vuorenmaa, J., Forsius, M. & Mannio, J. Increasing trends of total organic carbon concentrations in small forest lakes in Finland from 1987 to 2003. *Sci. Total Environ.* **365**, 47–65 (2006).
11. Siemens, J. The European carbon budget: a gap. *Science* **302**, 1681 (2003).
12. Evans, C. D. & Monteith, D. T. Chemical trends at lakes and streams in the UK Acid Waters Monitoring Network, 1988–2000: Evidence for recent recovery at a national scale. *Hydrol. Earth Syst. Sci.* **5**, 351–366 (2001).
13. Alewell, C., Mitchell, M. J., Likens, G. E. & Krouse, H. R. Sources of stream sulfate at the Hubbard Brook Experimental Forest. *Biogeochemistry* **44**, 281–299 (1999).

14. Tranvik, I. J. & Jansson, M. Terrestrial export of organic carbon. *Nature* **415**, 861–862 (2002).
15. Hejzlar, J., Dubrovsky, M., Buchtele, J. & Ruzicka, M. The apparent and potential effects of climate change on the inferred concentration of dissolved organic matter in a temperate stream (the Malse River, South Bohemia). *Sci. Total Environ.* **310**, 142–152 (2003).
16. Evans, C. D., Monteith, D. T. & Cooper, D. M. Long-term increases in surface water dissolved organic carbon: Observations, possible causes and environmental impacts. *Environ. Pollut.* **137**, 55–71 (2005).
17. Zhu, Y. & Day, R. L. Analysis of streamflow trends and the effects of climate in Pennsylvania, 1971 to 2001. *J. Am. Water Resour. Assoc.* **41**, 1393–1405 (2005).
18. de Wit, H. A., Hindar, A. & Hole, L. Winter climate affects long-term trends in stream water nitrate in acid-sensitive catchments in southern Norway. *Hydrol. Earth Syst. Sci.* **4**, 3055–3085 (2007).
19. Whitfield, P. H. & Cannon, A. J. Recent variations in climate and hydrology in Canada. *Can. Water Resour. J.* **25**, 19–65 (2000).
20. Hyvärinen, V. Trends and characteristics of hydrological time series in Finland. *Nordic Hydrol.* **34**, 71–90 (2003).
21. Wright, R. F. et al. Trends in nitrogen deposition and leaching in acid-sensitive streams in Europe. *Hydrol. Earth Syst. Sci.* **5**, 299–310 (2001).
22. Krug, E. C. & Frink, C. R. Acid rain on acid soil: a new perspective. *Science* **221**, 520–525 (1983).
23. Clark, J. M., Chapman, P. J., Heathwaite, A. L. & Adamson, J. K. Suppression of dissolved organic carbon by sulphate induced acidification during simulated droughts. *Environ. Sci. Technol.* **40**, 1776–1783 (2006).
24. de Wit, H. A., Groseth, T. & Mulder, J. Predicting aluminum and soil organic matter solubility using the mechanistic equilibrium model WHAM. *Soil Sci. Soc. Am. J.* **65**, 1089–1100 (2001).
25. Vogt, R. D., Rannekleiv, S. B. & Mykkelbost, T. C. The impact of acid treatment on soilwater chemistry at the HUMEX site. *Environ. Int.* **3**, 277–286 (1994).
26. Tipping, E. & Hurley, M. A. A model of solid-solution interactions in acid organic soils, based on the complexation properties of humic substances. *J. Soil Sci.* **39**, 505–519 (1988).
27. Tipping, E. & Woof, C. The distribution of humic substances between the solid and aqueous phases of acid organic soils; a description based on humic heterogeneity and charge-dependent sorption equilibria. *J. Soil Sci.* **42**, 437–448 (1991).
28. Evans, A. Jr, Zelazny, L. W. & Zipper, C. E. Solution parameters influencing dissolved organic carbon levels in three forest soils. *Soil Sci. Soc. Am. J.* **52**, 1789–1792 (1988).
29. Sokal, R. R. & Rohlf, F. J. *Biometry* 532–538 (W. H. Freeman, San Francisco, 1969).
30. Sen, P. K. On a class of aligned rank order tests in two-way layouts. *Ann. Math. Stat.* **39**, 1115–1124 (1968).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the LRTAP Working Group on Effects and the EU 6th Framework Programme Euro-limpacs for support in the production and analysis of international, quality-controlled, comparable data. We also acknowledge the work of the ICP Waters Programme Centre at the Norwegian Institute of Water Research (NIVA), where the data were collated, verified and archived. The authors are indebted to many colleagues and organisations who provided data for this assessment, including: T. A. Clair, S. Couture, C. Gagnon, D. K. McNicol, R. C. Weeber, A. Paterson (Canada); J. S. Kahl, J. Kellogg, K. Roy, M. R. Hale, D. R. DeWalle (USA); the Finnish Environment Institute (SYKE) and Regional Environmental Centres; the Norwegian Institute of Water Research (NIVA); the Swedish Environmental Protection Agency (Naturvårdsverket); and the UK Acid Waters Monitoring Network (supported by the Department for Environment Food and Rural Affairs) and supporting laboratories at: Fisheries Research Services, Pitlochry; Centre for Ecology and Hydrology, Wallingford; and the Environment Agency Llanelli. The information in this document has been funded in part by the US Environmental Protection Agency. It has been subjected to review by the National Health and Environmental Effects Research Laboratory, and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Author Contributions D.T.M. and J.L.S. formulated the working hypothesis and analysed and interpreted the trend data. C.D.E. and H.A.d.W. contributed to the development of the hypothesis, assisted in the interpretation of the data, provided additional text and edited the manuscript. M.F., T.H., A.W., B.L.S., D.S.J., B.K. and J. Vuorenmaa provided data and commented on the text. J.K. provided advice and ideas on processes and contributed data from the Czech Republic (not included in the final analysis due to number and length of time series) consistent with the hypothesis. J. Vesely contributed advice, ideas and data in the early stages of development of our work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.T.M. (dmonteit@geog.ucl.ac.uk).

METHODS

Data were drawn from a number of monitoring programmes with varying sampling protocols. Data for streams were accepted only if these represented multiple samples per year, while single annual samples for lakes were included provided these were collected in the same season for individual lakes. Analysis was restricted to sites with, at a minimum, annual measurements for ten of the 15 years between 1990 and 2004. Data for sites with more frequent sampling were reduced to one representative observation per year, by estimating annual mean values.

Given the multiple sources, all data were subjected to rigorous quality assurance. In addition to standard quality assurance procedures (for example, charge balance, calculated versus measured conductivity), we also examined the data for evidence of road salt inputs (by comparing Cl^- concentration to distance from the coast), or catchment sources of SO_4^{2-} , either of which could produce anomalous trends. We eliminated a handful of sites ($n = 15$) for which the SO_4^{2-} trends could not be attributed to changing deposition (that is, the surface water ΔSO_4^{2-} was more than three times the estimated deposition ΔSO_4^{2-}) and data from sites on older soils (for example, the southern Appalachian region of the US) where high levels of SO_4^{2-} adsorption lead to non-conservative behaviour. We also restricted our analysis to sites where DOC concentrations were sufficient to allow reliable quantification of trends; sites with median DOC concentrations of $<1 \text{ mg l}^{-1}$ were excluded from our analyses. Our final data set comprised estimates of trends and mean concentrations for chemical and air temperature parameters for 522 sites.

To determine the relative effects of anthropogenic and marine deposition factors on $\%\Delta\text{DOC}$ in our model, we first estimated the change in non-marine and marine sulphate (that is, ΔxSO_4^{2-} and ΔmSO_4^{2-}) at each site using a sea-salt-correction factor, so that: $\Delta\text{mSO}_4^{2-} = \Delta\text{Cl}^- \times 0.104$ and $\Delta\text{xSO}_4^{2-} = \Delta\text{SO}_4^{2-} - \Delta\text{mSO}_4^{2-}$. The effects of changing anthropogenic and marine anions on DOC were then determined from the model coefficients:

$$\%\Delta\text{DOC (anthropogenic)} = -0.557\Delta\text{xSO}_4^{2-}$$

$$\%\Delta\text{DOC (marine induced)} = -0.557\Delta\text{mSO}_4^{2-} - 0.504\Delta\text{Cl}^- + 0.0033\Delta\text{Cl}^- \times \text{CaMg}^*$$

Results, in the form of mean effects on $\%\Delta\text{DOC}$ per region, are provided in Supplementary Table 5.

Onset of submarine debris flow deposition far from original giant landslide

P. J. Talling¹, R. B. Wynn², D. G. Masson², M. Frenz², B. T. Cronin³, R. Schiebel², A. M. Akhmetzhanov², S. Dallmeier-Tiessen⁴, S. Benetti², P. P. E. Weaver², A. Georgiopoulou², C. Zühlsdorff⁴ & L. A. Amy¹

Submarine landslides can generate sediment-laden flows whose scale is impressive. Individual flow deposits have been mapped that extend for 1,500 km offshore from northwest Africa^{1–7}. These are the longest run-out sediment density flow deposits yet documented on Earth. This contribution analyses one of these deposits, which contains ten times the mass of sediment transported annually by all of the world's rivers⁸. Understanding how this type of submarine flow evolves is a significant problem, because they are extremely difficult to monitor directly⁹. Previous work has shown how progressive disintegration of landslide blocks can generate debris flow, the deposit of which extends downslope from the original landslide^{10–13}. We provide evidence that submarine flows can produce giant debris flow deposits that start several hundred kilometres from the original landslide, encased within deposits of a more dilute flow type called turbidity current. Very little sediment was deposited across the intervening large expanse of sea floor, where the flow was locally very erosive. Sediment deposition was finally triggered by a remarkably small but abrupt decrease in sea-floor gradient from 0.05° to 0.01°. This debris flow was probably generated by flow transformation from the decelerating turbidity current. The alternative is that non-channelized debris flow left almost no trace of its passage across one hundred kilometres of flat (0.2° to 0.05°) sea floor. Our work shows that initially well-mixed and highly erosive submarine flows can produce extensive debris flow deposits beyond subtle slope breaks located far out in the deep ocean.

Giant submarine landslides can generate rapidly flowing mixtures of sediment and water. The seminal event for understanding such landslide-generated flows occurred offshore from the Grand Banks, Newfoundland, in 1929 (ref. 10). A large earthquake caused landslides across 100 km of continental slope. A series of submarine cables were then broken downslope from the landslides. These breaks recorded a flow, the frontal velocity of which was 19 m s⁻¹ on a gradient of only 0.25°. The Grand Banks event showed that a single submarine landslide can generate two distinct types of sediment-laden flow¹⁰, termed 'turbidity current' and 'debris flow'. Turbidity currents are fully turbulent and relatively dilute. We adopt the most common definition of turbidity currents, as flows in which sediment is supported primarily by fluid turbulence¹⁴. Debris flows have a higher sediment concentration that suppresses turbulence, such that other processes become more important than fluid turbulence for supporting sediment^{14,15}. These processes include grain-to-grain collisions, reduced excess particle density, and yield strength of the sediment–water mixture¹⁵. Debris flows defined in this way are typically laminar, but can be weakly turbulent^{15,16}.

Previous studies, including those of the Grand Banks event, have shown that submarine debris flows can be generated by progressive disintegration of strata within an initial large submarine landslide^{9–12}.

Debris flows formed in this way extend downslope from the landslide, sometimes for several hundred kilometres, and contain intact blocks of original landslide material. This type of debris flow deposit is typically tens of metres thick^{9,10,12}. Here we describe a different type of large-scale debris flow deposit, and propose a new mechanism to explain its origin. This debris-flow deposit is unusual because it starts several hundred kilometres from the original landslide, is less than two metres thick, lacks any intact blocks of landslide material, and is encased within a turbidity-current deposit formed during the same event.

Our study is based on an analysis of shallow sediment cores from the Agadir basin, and the Seine and Madeira abyssal plains, located offshore from northwest Africa^{1–7} (Fig. 1). These cores contain a sequence of deposits spanning the last ~200 thousand years (kyr) (ref. 6). The majority of flows originated from two distinct sources: the Moroccan continental margin and the volcanic Canary Islands^{3,6} (Fig. 1). Flow deposits were correlated between cores (Supplementary Fig. 1). This correlation framework is unusually detailed and extensive⁶, and allows us to analyse how the deposit from each flow varies spatially.

Flow deposits typically comprise sand overlain by mud. The first type of sand contains a small fraction of interstitial mud (sediment <32 µm is <7%), becomes finer upwards, and has ubiquitous planar lamination or ripple cross-lamination (Fig. 2 and Supplementary Fig. 3). There is a gradual transition from this type of sand into overlying mud. These features indicate deposition from a low-sediment-concentration suspension inferred to be a turbidity current^{6,14}. The second type of sand interval occurs in only one flow deposit, which we call bed 5 (refs 6, 17). This type of sand has four distinct features: it is ungraded, there are no laminations, it contains a much higher percentage of mud (sediment <32 µm is >30%), and it is separated from overlying mud by an abrupt change in grain size (Fig. 2; Supplementary Fig. 3). These four features indicate sand deposition en masse from a higher-concentration mud-rich sediment mixture¹⁸ inferred to be a cohesive debris flow^{11,17}.

A turbidite (the deposit formed by a turbidity current) wholly encases the debrite (the deposit formed by a debris flow) within bed 5, indicating that both were deposited during the same flow event. Background oceanic sedimentation deposited hemipelagic sediment between other flow deposits (Supplementary Fig. 1), but hemipelagic sediment is absent between the turbidite and the debrite in bed 5.

The flow that deposited bed 5 contained ~125 km³ of sediment⁶ (22.5 × 10¹³ kg of sediment, assuming a density of 1,800 kg m⁻³). Some submarine density flows originate from flood discharges from rivers¹⁴. However, even during very large historical floods, rivers supplied only 10⁹ to 10¹¹ kg of sediment to the ocean¹⁹. The annual flux of sediment from all of the world's rivers to the oceans¹³ is ~2 × 10¹³ kg. The large volume of bed 5 indicates flow initiation by a submarine landslide.

¹Department of Earth Sciences, University of Bristol, Queens Road, Bristol, BS8 1RJ, UK. ²National Oceanography Centre Southampton, European Way, Southampton, SO14 3ZH, UK. ³School of Geosciences, University of Aberdeen, St Mary's, Elphinstone Road, Aberdeen, AB24 3UE, UK. ⁴Department of Sedimentology and Paleooceanography, University of Bremen, Klagenfurter Strasse, 28334 Bremen, Germany.

Debrite and turbidite intervals within bed 5 have similar compositions rich in shallow-water shelf sediment⁶ (Supplementary Fig. 3). This sediment composition, together with spatial trends in grain size, indicates that the original landslide occurred on the Moroccan continental margin near the Agadir canyon^{2–6}. Seismic reflection profiles show that shallow landslides and debris flows are widespread along the flanks of this canyon, although the landslide that generated bed 5 cannot be identified using available data. The debrite in bed 5 is found in the centre of the Agadir basin, several hundred kilometres from the nearest plausible location for this landslide.

The event responsible for bed 5 probably caused significant erosion in the lower Agadir canyon, although preceding flows could also have contributed to this erosion. Our cores suggest that up to 4 m of hemipelagic sediment was eroded in locations 200–300 m above the canyon floor (Supplementary Fig. 2). Bed 5 directly overlies these erosional surfaces, where it comprises only a few centimetres of fine turbidite sand or mud. A spectacular field of ~10-m-deep scours occurs at the canyon mouth⁷ (Supplementary Fig. 2). This scour field was probably eroded primarily by the flow responsible for bed 5, because this flow was the coarsest-grained and presumably the most powerful flow in the last 200 kyr (ref. 6).

Beyond the canyon mouth there is a ~100-km-long area of open sea floor that we term the 'exit ramp' (Figs 1 and 2). The flow deposited very little sediment on the exit ramp and locally eroded up to one metre of underlying material (Supplementary Fig. 1a). An abrupt change in gradient from 0.05° to 0.01° separates the exit ramp from the Agadir basin plain. This remarkably small change in gradient coincides with the start of debrite deposition, which extends for a further ~250 km (Fig. 2). The debrite is underlain by turbidite sand in the distal Agadir basin, presumably because the turbidity current outran the debris flow.

The turbidity current spread more widely than the debris flow (Supplementary Figs 4 and 5) and continued into the Madeira abyssal plain^{2–6}.

Simple calculations indicate that flow was at least weakly turbulent on the exit ramp. The debrite within bed 5 is unusually thin, suggesting that this debris flow had particularly low yield strength. A yield strength of approximately 1.3 Pa is estimated for motion of a ~3-m-thick debris flow on a gradient of ~0.01° (assuming Bingham plastic rheology and a flow density of <1,250 kg m⁻³)²⁰. Debris flows with such low yield strength can support sand²¹ but are predicted to become turbulent when their velocity exceeds ~0.8 m s⁻¹ (ref. 20). This velocity is likely to be exceeded on gradients steeper than ~0.02°, assuming a friction coefficient²² of ~0.004. Similar benthic foraminiferal assemblages in the turbidite and the debrite also suggest that the flow was originally well-mixed, and therefore at least weakly turbulent (Supplementary Discussion).

Two mechanisms for debris flow formation are evaluated. In the first model, the debris flow originates by means of a complete disintegration of the original landslide, but deposited no sediment for ~200 km across the canyon mouth and exit ramp (Fig. 3c). This debris flow could have been weakly turbulent¹⁹, but turbulence was not the primary sediment support mechanism^{17,18}. In the second model, debris flow originates from a decelerating turbidity current generated by the initial landslide (Fig. 3b). The models differ mainly in whether turbulence is the primary mechanism for sediment support during flow bypass across the lower canyon and exit ramp. We cannot determine unequivocally which support mechanism was most important in this part of the flow. Lack of debris-flow deposits and evidence for fast-flow velocities (giant scours and minimal deposition, for example) indicate that sediment was probably supported primarily by turbulence.

We propose that debris-flow deposition was probably triggered by rapid deceleration of this initial (weakly or strongly turbulent) sediment

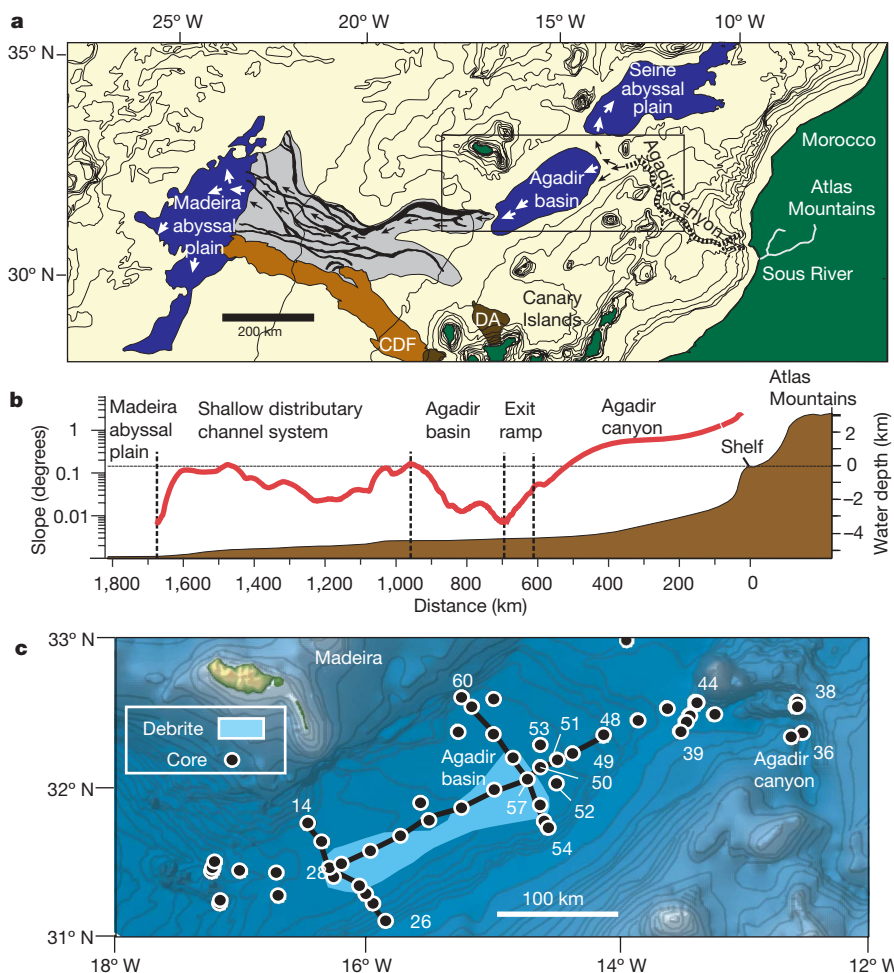


Figure 1 | Location of main features in study area offshore from northwest Africa. **a**, The Agadir canyon (stippled), Agadir basin, Seine and Madeira abyssal plains; channel network between Agadir basin and Madeira abyssal plain (grey shade); Canary debris flow (CDF; shaded brown); debris avalanches (DA) from Canary Islands^{1–7}. The path of the flow that deposited bed 5 is shown by arrows. Box indicates area shown in **c**. **b**, Bathymetric contours spaced at 500 m intervals. **b**, Change in seafloor gradient (red line) plotted against distance along flow path. **c**, Location of cores (filled circles), debris-flow deposit (debrite), and the location of the cross-sections shown in Fig. 2 (28–48) and in Supplementary Figs 4 and 5 (the three bold black lines) in Agadir basin.

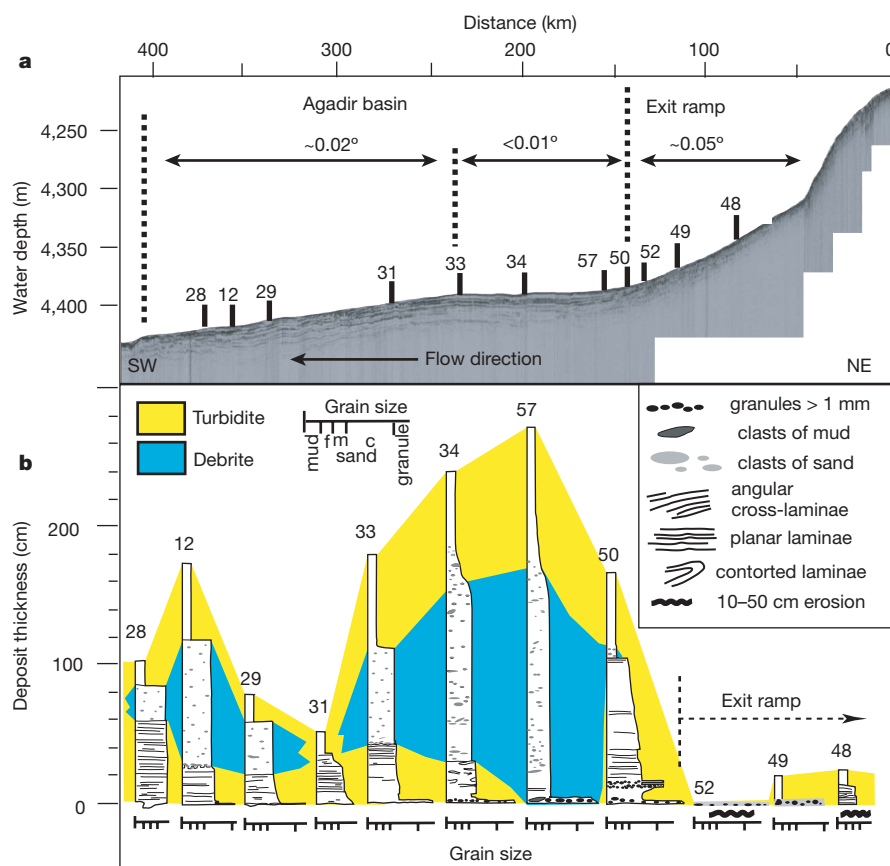


Figure 2 | Change in seafloor gradient and the shape of bed 5. **a**, High-resolution (3.5 kHz) seismic reflection profile along the exit ramp and axis of the Agadir basin. **b**, Sedimentary logs documenting the thickness, grain size and sedimentary structures within bed 5. Intervals deposited by turbidity current and debris flow are indicated. Numbered core locations are shown in Fig. 1c. Hemipelagic intervals occur above and below bed 5 (Supplementary Fig. 1). f, fine; m, medium; c, coarse.

suspension (Fig. 3). Rapid flow deceleration increased the gradient Richardson number significantly and thus reduced turbulent mixing. Sand settled more rapidly from the flow once turbulent mixing was suppressed²³. Turbidity currents have a zone of minimum turbulence

near their velocity maximum, which could have helped to trap sediment near the bed²³. A denser sediment suspension thus formed near the sea floor, overlain by an abrupt decrease in sediment concentration, or lutocline^{1,24,25}. During the latter stages of this process, cohesive mud

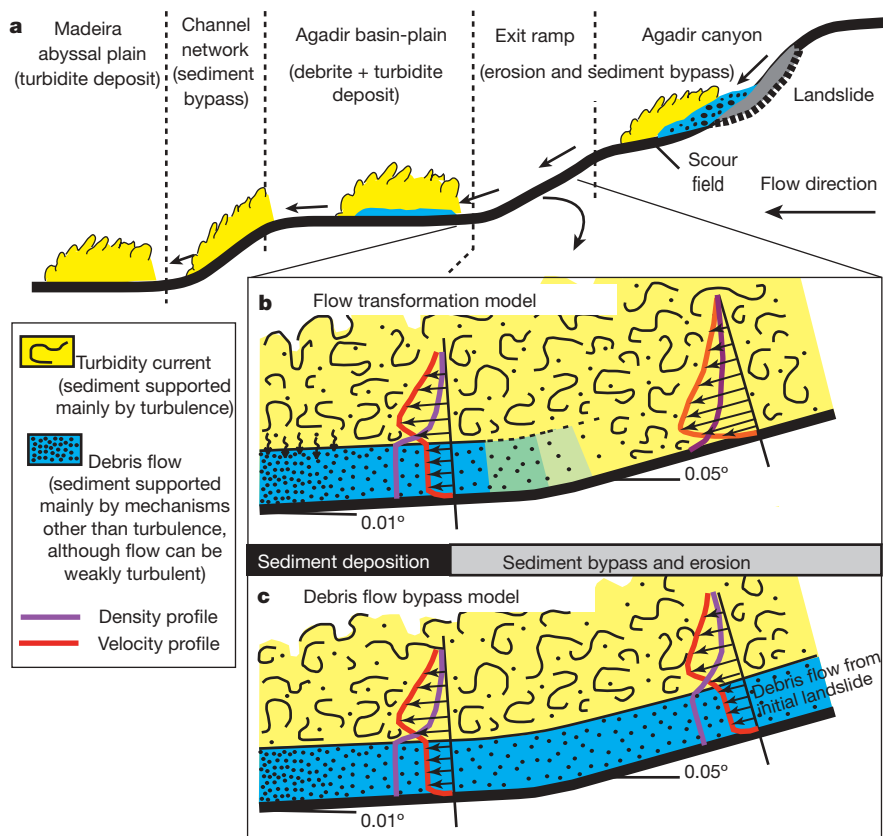


Figure 3 | Evolution of flow event that deposited bed 5 showing two alternative mechanisms for generating the debris flow. **a**, Evolution of entire flow event from Agadir canyon to Madeira abyssal plain. **b**, Debris flow forms owing to flow transformation from turbidity current beyond break in slope. The event comprises only well-mixed turbidity current in the lower Agadir canyon and exit ramp. **c**, Debris flow forms by disintegration of initial landslide in upper canyon. The debris flow is present in the lower canyon and exit ramp but leaves no deposit (bypasses). We note that the turbidity current is actually much thicker than debris flow in both models. Greenish shaded areas indicate gradual transition from turbidity current to debris flow.

within the suspension developed sufficient yield strength to damp turbulence further²³. Eventually, as turbulence was fully suppressed, yield strength replaced turbulence as the mechanism that supported sand within the flow²³. En masse settling then generated the ungraded, structureless, mud-rich and sharp-topped debrite sand¹⁸.

This transformation was probably triggered by deceleration near the slope break at the base of the exit ramp, because thick debrite starts just beyond this point. Flow transformation could have initiated on the exit ramp (as recorded by thin granule layers with mud matrix in cores 52 and 48; Fig. 3), or across a wider expanse of the Agadir basin. Seafloor erosion in the lower canyon probably increased sediment concentration, and 'primed' this particular flow to transform. This would explain why other large-volume flow deposits⁶ do not contain debrites near the same slope break (Supplementary Fig. 1).

The origin of sandy debris flows is important not just for understanding how sediment is delivered to the deep ocean by huge flows, but also has potential implications for hydrocarbon exploration. This is because some of the world's largest oil and gas reserves^{26–28} occur within ancient rock sequences that contain similar debrites, deposited during the same event as surrounding turbidites¹⁷. Mud-rich debrite sandstone is a barrier to subsurface flow of oil and gas within more permeable turbidite sandstone. It is necessary to predict the extent and shape of these debrite sandstones to recover hydrocarbon reserves efficiently. Our work shows that extensive debrites can form down flow from abrupt slope breaks and areas of significant seafloor erosion. The 'linked' debrite forms the centre of the deposit and is encased within turbidite sandstone and mudstone.

We also demonstrate that voluminous submarine flows can leave behind remarkably little sediment on low slopes (0.2 to 0.05° in our case). This result also has important implications for assessment of submarine geohazards. Exploration for oil and gas now commonly targets 'deep-water' oceanic settings on the continental slope^{26–28}. Large-volume density flows are potentially catastrophic for seafloor installations involved in oil and gas recovery^{26,27}, whose value may be several hundred million dollars²⁸. These installations are typically sited on slopes >0.05°. Cores collected next to the installations for geotechnical purposes are often used for subsequent geohazard analysis. Our work suggests that such cores could contain a subtle record of very large volume flows. The clearest record of these flows is found by coring low-gradient basin plains, downslope from installations. The frequency of large-scale flows and their parent landslides could be greatly underestimated if cores from only higher-gradient settings next to installations are used.

METHODS SUMMARY

Individual flow deposits were correlated between cores using sand-fraction composition, the ratio of coccolith species within mud, magnetic susceptibility, colour, thickness, and relative position of layers within the core^{3,6} (for example, Supplementary Fig. 1). Ratios of coccolith nannofossil species within hemipelagic mud were used to date the flows^{2–6}, together with an oxygen isotope stratigraphy. We analyse the deposit of a particular flow (bed 5) that occurred ~60 kyr ago⁶. Coccolith and benthic foraminifera mixtures are documented within bed 5. Coccolith nannofossils record the relative age of sediment incorporated into the flow^{2–4,6}. Older coccoliths originate from more deeply buried sediment layers. Benthic foraminiferal microfossil assemblages record the water depth from which these sand-sized particles originated^{29,30}.

Received 11 January; accepted 24 September 2007.

- McCave, I. N. & Jones, K. P. N. Deposition of ungraded muds from high-density non-turbulent turbidity currents. *Nature* **133**, 250–252 (1988).
- Weaver, P. P. E. & Thomson, J. Calculating erosion by deep-sea turbidity currents during initiation and flow. *Nature* **364**, 136–138 (1993).
- Weaver, P. P. E., Rothwell, R. G., Ebbing, J., Gunn, D. E. & Hunter, P. M. Correlation, frequency of emplacement and source directions of megaturbidites on the Madeira Abyssal Plain. *Mar. Geol.* **109**, 1–20 (1992).
- Weaver, P. P. E. Determination of turbidity current erosional characteristics from re-worked coccolith assemblages, Canary Basin, Northeast Atlantic. *Sedimentology* **41**, 1025–1038 (1994).
- Masson, D. G. Late Quaternary turbidity current pathways to the Madeira Abyssal Plain and some constraints on turbidity current mechanisms. *Basin Res.* **6**, 17–33 (1994).

- Wynn, R. B., Weaver, P. P. E., Masson, D. G. & Stow, D. A. V. Turbidite depositional architecture across three inter-connected deep-water basins on the Northwest African Margin. *Sedimentology* **49**, 669–695 (2002).
- Wynn, R. B., Kenyon, N. H., Masson, D. G., Stow, D. A. V. & Weaver, P. P. E. Characterization and recognition of deep-water channel-lobe transition zones. *Am. Assoc. Petrol. Geol. Bull.* **86**, 1441–1462 (2002).
- Khrifounoff, A. et al. Direct observation of intense turbidity current activity in the Zaire Submarine Valley at 4000 m water depth. *Mar. Geol.* **194**, 151–158 (2003).
- Masson, D. G., Harbitz, C. B., Wynn, R. B., Pedersen, G. & Lovholt, F. Submarine landslides: processes, triggers and hazard prediction. *Phil. Trans. R. Soc. A.* **364**, 2009–2039 (2006).
- Piper, D. J. W., Cochonat, P. & Morrison, M. L. The sequence of events around the epicentre of the 1929 Grand Banks earthquake: initiation of the debris flows and turbidity current inferred from side scan sonar. *Sedimentology* **46**, 79–97 (1999).
- Mulder, T., Savoye, B. & Syvitski, J. P. M. Numerical modelling of a mid-sized gravity flow: the 1979 Nice turbidity current (dynamics, processes, sediment budget and seafloor impact). *Sedimentology* **44**, 305–326 (1997).
- Gee, M. J. R., Masson, D. G., Watts, A. B. & Allen, P. A. The Saharan debris flow: an insight into the mechanics of long runout submarine debris flow. *Sedimentology* **46**, 317–335 (1999).
- Milliman, J. D. & Syvitski, J. P. M. Geomorphic tectonic controls of sediment discharge to the ocean – the importance of small mountain rivers. *J. Geol.* **100**, 525–544 (1992).
- Mulder, T. & Alexander, J. The physical character of subaqueous sedimentary density flows and their deposits. *Sedimentology* **48**, 269–299 (2001).
- Iverson, R. M. The physics of debris flows. *Rev. Geophys.* **35**, 245–296 (1997).
- Costa, J. E. & Williams, G. P. *Debris-Flow Dynamics*. [22 minute video] USGS Open File Rep. 84–606 (1984).
- Talling, P. J., Amy, L. A., Wynn, R. B., Peakall, J. & Robinson, M. Beds comprising debrite sandwiched within co-genetic turbidite: origin and widespread occurrence in distal depositional environments. *Sedimentology* **51**, 163–194 (2004).
- Amy, L. A., Talling, P. J., Edmonds, V. O., Sumner, E. J. & Leseur, A. An experimental investigation on sand-mud suspension settling behaviour and implications for bimodal mud content of submarine flow deposits. *Sedimentology* **53**, 1411–1434 (2006).
- Dadson, S. et al. Hyperpycnal river flows from an active mountain belt. *J. Geophys. Res.* **110**, doi:10.1029/2004JF000244 F04016 (2005).
- Schwab, W. C. et al. Sediment mass-flow processes on a depositional lobe, Outer Mississippi Fan. *J. Sedim. Res.* **66**, 916–927 (1996).
- Hampton, M. A. The role of subaqueous debris flow in generating turbidity currents. *J. Sedim. Petrol.* **42**, 775–793 (1972).
- Bowen, A. J., Normark, W. B. & Piper, D. J. W. Modeling of turbidity currents on Navy Submarine Fan, California continental borderland. *Sedimentology* **31**, 169–185 (1984).
- Baas, J. H. & Best, J. L. Turbulence modulation in clay-rich sediment-laden flows and some implications for sediment deposition. *J. Sedim. Res.* **72**, 336–340 (2002).
- Trowbridge, J. H. & Kineke, G. C. Structure and dynamics of fluid muds on the Amazon continental shelf. *J. Geophys. Res.* **99**, 865–874 (1994).
- Winterwerp, J. C. Stratification effects by fine suspended sediment at low, medium, and very high concentrations. *J. Geophys. Res.* **111**, doi:10.1029/2005JC003019 C05012 (2006).
- Barley, B. Deepwater problems around the world. *Leading Edge* **18**, 488–494 (1999).
- Sparkman, G. An introduction to special section on deepwater. *Leading Edge* **19**, 381 (2000).
- Eyton, D. The journey to deepwater operatorship. *Leading Edge* **24**, 902–904 (2005).
- Lutze, G. F. & Coulbourn, W. T. Recent benthic foraminifera from the continental margin of Northwest Africa: Community structure and distribution. *Mar. Micropaleo.* **8**, 361–401 (1983/4).
- Kuhlmann, H., Freudenthal, T., Helmke, P. & Meggers, H. Reconstruction of paleoceanography off NW Africa during the last 40,000 years: influence of local and regional factors on sediment accumulation. *Mar. Geol.* **207**, 209–224 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the officers and crew of RSS *Charles Darwin* and Shipboard Scientific Party and Coring Technical Team on cruise CD166. This work was supported by the UK Natural Environment Research Council, ConocoPhillips, BHP Billiton, ExxonMobil and Shell.

Author Contributions P.J.T. wrote the manuscript, incorporating comments from all co-authors. R.B.W. led the research project and was chief scientist on cruise CD166. B.T.C. was co-chief scientist during cruise CD166. P.J.T., D.G.M., M.F., A.M.A., S.D.-T., A.G. and C.Z. participated in data collection during this cruise. P.J.T., R.B.W. and D.G.M. visually described cores. M.F. completed grain size analyses. R.S. analysed benthic foraminifera assemblages. A.M.A. processed bathymetric and shallow seismic data. S.D.-T., assisted by C.Z., quantified sediment composition and imaged sedimentary structures. S.B., P.P.E.W. and A.G. analysed coccolith assemblages. L.A.A. contributed to data interpretation. R.B.W., P.J.T. and B.T.C. obtained industry funding.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.J.T. (peter.talling@bris.ac.uk).

Enhanced biological carbon consumption in a high CO₂ ocean

U. Riebesell¹, K. G. Schulz¹, R. G. J. Bellerby^{2,3}, M. Botros¹, P. Fritsche¹, M. Meyerhöfer¹, C. Neill², G. Nondal^{2,3}, A. Oschlies¹, J. Wohlers¹ & E. Zöllner¹

The oceans have absorbed nearly half of the fossil-fuel carbon dioxide (CO₂) emitted into the atmosphere since pre-industrial times¹, causing a measurable reduction in seawater pH and carbonate saturation². If CO₂ emissions continue to rise at current rates, upper-ocean pH will decrease to levels lower than have existed for tens of millions of years and, critically, at a rate of change 100 times greater than at any time over this period³. Recent studies have shown effects of ocean acidification on a variety of marine life forms, in particular calcifying organisms^{4–6}. Consequences at the community to ecosystem level, in contrast, are largely unknown. Here we show that dissolved inorganic carbon consumption of a natural plankton community maintained in mesocosm enclosures at initial CO₂ partial pressures of 350, 700 and 1,050 µatm increases with rising CO₂. The community consumed up to 39% more dissolved inorganic carbon at increased CO₂ partial pressures compared to present levels, whereas nutrient uptake remained the same. The stoichiometry of carbon to nitrogen drawdown increased from 6.0 at low CO₂ to 8.0 at high CO₂, thus exceeding the Redfield carbon:nitrogen ratio of 6.6 in today's ocean⁷. This excess carbon consumption was associated with higher loss of organic carbon from the upper layer of the stratified mesocosms. If applicable to the natural environment, the observed responses have implications for a variety of marine biological and biogeochemical processes, and underscore the importance of biologically driven feedbacks in the ocean to global change.

Throughout Earth's history, the ocean has had a crucial role in modulating atmospheric carbon dioxide through a variety of physical, chemical and biological processes. The same processes are involved in the ocean's response to anthropogenic perturbations of the global carbon cycle. A key process responsible for about three-quarters of the surface to deep-ocean gradient in dissolved inorganic carbon (DIC) is the biological carbon pump⁸. This transports carbon bound by photosynthesis from the sunlit surface layer to the deep ocean. Integrated over the global ocean, the biotically driven surface to deep-ocean DIC gradient of ~220 µmol kg⁻¹ (ref. 9) amounts to ~2,500 petagrams of carbon (Pg C) (1 Pg = 10¹⁵ g), that is, 3.5 times the atmospheric carbon pool. Small changes in this pool, for example, caused by biological responses to ocean change, would have a strong affect on atmospheric CO₂.

At present, one of the most far-reaching global perturbations of the marine environment is caused by the massive invasion of fossil fuel CO₂ into the ocean, making it the second largest sink for anthropogenic carbon dioxide after the atmosphere itself. CO₂ entering the ocean alters the seawater carbonate equilibrium, decreasing pH and shifting dissolved inorganic carbon away from carbonate (CO₃²⁻) towards more bicarbonate (HCO₃⁻) and CO₂. For a 'business-as-usual' emission scenario in the year 2100 (IS92a), the CO₂ concentration will rise

by about a factor of two relative to the present value (380 µatm), and could increase by a factor of three by the middle of the next century¹⁰. This will cause seawater pH to further drop by 0.3 and 0.6 pH units, respectively, in addition to the 0.12 pH-unit decrease that has occurred since pre-industrial times³. Changes in seawater chemistry of this magnitude are expected to have adverse effects, not only on individual species but also at the community and ecosystem level¹¹.

To investigate the effect of rising CO₂ on a natural plankton community, we conducted a mesocosm CO₂ perturbation study in the Raune Fjord in southern Norway. Nine enclosures, each containing 27 m³ of ambient water, were aerated with CO₂-enriched air to achieve concentrations of 350 µatm (1 × CO₂), 700 µatm (2 × CO₂) and 1,050 µatm (3 × CO₂). After nutrient addition, the development and decline of a plankton bloom was monitored over 24 days.

A rapid decline in CO₂ partial pressures (*p*CO₂) owing to photosynthetic carbon fixation led to a minimum *p*CO₂ at day 12, at which point nitrate concentrations were close to exhaustion (Fig. 1). The greater drop in *p*CO₂ at increased CO₂ levels was caused by the lower buffer capacity of sea water under these conditions. No difference between CO₂ treatments was observed in the drawdown of nitrate (Fig. 1c), phosphate and silicate (see Methods, Supplementary Fig. 1). Phytoplankton biomass, depicted by chlorophyll *a* concentrations (Fig. 1b), peaked on day 10. This corresponded to the maximum in particulate organic carbon (Methods, Supplementary Fig. 2) and coincided with phosphate exhaustion. The phytoplankton bloom was initially dominated by diatoms, which reached a maximum 1–2 days before the peak of the bloom owing to silicate limitation on day 9 (Fig. 2). This was followed by a dominance of the prymnesiophytes, primarily the coccolithophore *Emiliania huxleyi*, which had maximum cell abundances of (3.5–6.2) × 10⁶ cells l⁻¹ on days 10–11. Whereas diatoms and prymnesiophytes accounted for 85–90% of phytoplankton biomass during the bloom, prasinophytes, dinoflagellates and cyanobacteria dominated after the decline of the bloom. Neither phytoplankton composition nor succession differed significantly between CO₂ treatments.

A distinct CO₂ treatment effect is evident for cumulative DIC drawdown owing to production of organic matter (ΔDIC_{org}, Fig. 3a), averaging 79, 100 and 110 µmol kg⁻¹ at the peak of the bloom (day 12) in water at 1 × CO₂, 2 × CO₂ and 3 × CO₂ concentration, respectively. Thus, for the same uptake of inorganic nutrients, net community carbon consumption under increased CO₂ exceeded present rates by 27% (2 × CO₂) and 39% (3 × CO₂). Continuous oxygen measurements in mesocosms 2 (3 × CO₂), 5 (2 × CO₂) and 8 (1 × CO₂), which revealed up to 20 µmol kg⁻¹ higher O₂ concentrations at increased CO₂ (data not shown), indicate enhanced net photosynthesis to be the source of the observed CO₂ effect. Possible causes for this include increased photosynthetic

¹Leibniz Institute of Marine Sciences, IFM-GEOMAR, 24105 Kiel, Germany. ²Bjerknes Centre for Climate Research, ³Geophysical Institute, University of Bergen, Allégaten 70, 5007 Bergen, Norway.

carbon fixation owing to lower energetic cost of carbon acquisition, reduced photorespiration owing to a higher $\text{CO}_2:\text{O}_2$ ratio, and decreased organic matter degradation, for example, owing to changes in food quality.

Whereas the cumulative carbon to nitrate (C:N) drawdown remained slightly below the Redfield C:N ratio of 6.6 under $1 \times \text{CO}_2$ (Fig. 3c), this ratio increased to 7.1 and 8.0 under $2 \times \text{CO}_2$ and $3 \times \text{CO}_2$, respectively. The ratio of particulate organic carbon to particulate organic nitrogen (POC/PON) accumulating in the surface layer, in contrast, closely adhered to the Redfield C:N ratio independent of CO_2 treatment (see Methods, Supplementary Fig. 2). Deviation of DIC/NO_3^- drawdown from POC/PON build-up, previously reported from both mesocosm and field studies^{12,13}, was attributed to dissolved organic carbon (DOC) release¹⁴. DOC concentrations also increased during bloom development in our study, but surface-layer DOC accumulations of $\sim 25 \mu\text{mol kg}^{-1}$ remained well below levels expected from the difference between DIC drawdown and POC build-up (see below). A possible fate of DOC is indicated by a fourfold increase in the concentration of transparent exopolymer particles (TEP), which are known to originate from dissolved precursors and to accelerate particle aggregation and sinking¹⁵.

The loss of organic carbon from the upper mixed layer (UML) through sinking is given by the difference between DIC drawdown and build-up of POC and DOC, $\Delta C_{\text{loss}} = \Delta \text{DIC}_{\text{org}} - (\Delta C_{\text{POC}} + \Delta C_{\text{DOC}})$. Between 15% and 30% of organically fixed carbon was lost from the UML daily, with consistently higher loss rates at increased CO_2 (Fig. 3b). The 12-day cumulative carbon loss at $3 \times \text{CO}_2$ exceeded that under $1 \times \text{CO}_2$ by about $30 \mu\text{mol kg}^{-1}$, which closely resembles the difference in net inorganic carbon drawdown between

the lowest and highest CO_2 treatment. Enhanced particle sinking at increased CO_2 may explain why POC and TEP concentrations did not show a significant difference between treatments. In fact, positive correlations between TEP production and CO_2 concentration were obtained in natural plankton assemblages dominated by diatoms and cyanobacteria¹⁶ and in batch culture incubations with *E. huxleyi*¹⁷.

The extent to which the observed CO_2 sensitivity in phytoplankton carbon consumption and stoichiometry in carbon-to-nutrient drawdown can be extrapolated to other marine ecosystems remains to be seen. In accordance with our results, enhanced carbon fixation of up to 15% in response to three-times increased CO_2 was observed during incubations of natural phytoplankton assemblages from the nutrient-poor central Atlantic¹⁸. Higher carbon-to-nitrate drawdown at increased CO_2 concentrations also occurred in a previous mesocosm CO_2 perturbation experiment during a bloom of *E. huxleyi*¹⁹. The concordant results indicate the existence of a widespread mechanism under high CO_2 conditions. The phytoplankton groups dominating in the mesocosm studies—diatoms and coccolithophores—are also the main primary producers in high productivity areas and are the principal drivers of biologically induced carbon export to the deep sea.

Apart from seawater acidification, the ocean in a high CO_2 world will experience other changes, including higher surface temperatures, enhanced stratification and decreased mixed-layer depths. Although the direct effect of rising temperatures on net community carbon-to-nutrient ratios is unclear, at the organism level most experimental

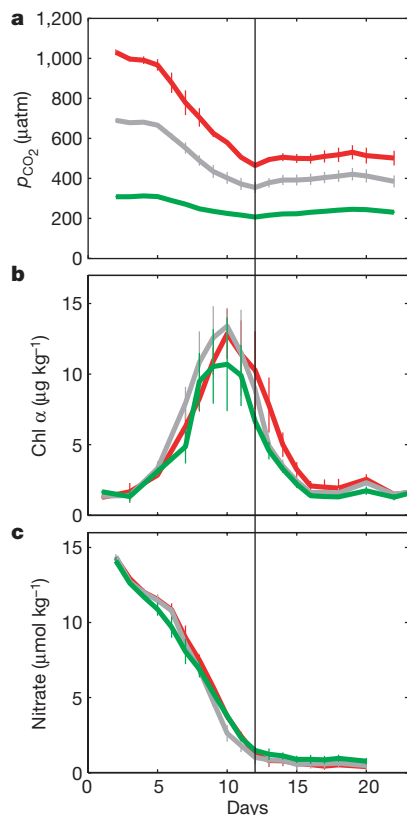


Figure 1 | Growth conditions in experimental mesocosms. **a**, CO_2 partial pressures (p_{CO_2}); **b**, chlorophyll *a* concentrations (Chl *a*); and **c**, nitrate concentrations in the upper 5.5 m mixed layer (UML) during the experiment. Values are means of triplicate CO_2 treatments with initial p_{CO_2} of 350 μatm (green), 700 μatm (grey) and 1,050 μatm (red). Error bars denote ± 1 standard deviation. The vertical line indicates time of deep water injection into UML owing to vertical mixing (day 12).

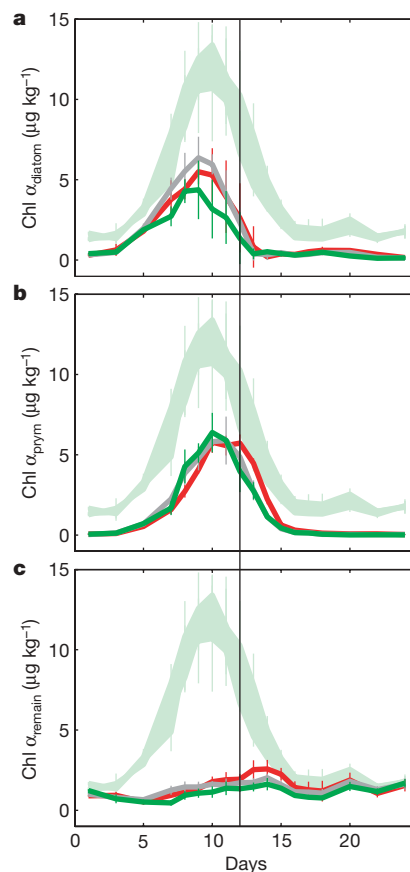


Figure 2 | Development of dominant phytoplankton groups. Chlorophyll *a* equivalents in the UML of: **a**, diatoms (Chl a_{diatom}); **b**, prymnesiophytes (Chl a_{pym}), predominantly the coccolithophore *E. huxleyi*; and **c**, remaining chlorophyll *a* (Chl a_{remain}), primarily containing prasinophytes, dinoflagellates and cyanobacteria. Group-specific chlorophyll *a* equivalents were derived from high-performance liquid chromatography pigment analysis using CHEMTAX algorithms (for details see Methods). The light-green curve represents the range of total chlorophyll *a* as depicted in Fig. 1b. Colour code is as in Fig. 1. The vertical line indicates time of deep water injection into UML owing to vertical mixing (day 12). Error bars denote ± 1 standard deviation.

studies indicate an increase in the C:N ratio with increasing temperature²⁰. Decreased mixed-layer depth tends to increase phytoplankton carbon-to-nutrient ratios²¹, which would augment the direct CO₂ effect on C:N:P stoichiometry observed in this study. Increased stratification, in contrast, would decrease the supply of nutrients to the surface layer, reducing overall primary and export production. Thus, concurrent with a decrease in the strength of the biological pump caused by a lower nutrient supply, the pump's efficiency is likely to increase at increased CO₂.

Shifting of the C:N:P stoichiometry of marine primary production has long been recognized as one of the most powerful mechanisms determining ocean-atmosphere carbon partitioning²². Carbon relative to nitrogen drawdown in excess of the Redfield C:N ratio, often referred to as carbon over-consumption²³, is reported for various oceanographic regimes^{12,13,23,24}, but was generally considered to be a transient phenomenon. Whether this also applies to the observed CO₂ sensitivity of plankton carbon consumption and to what extent the mechanism(s) underlying this trend are applicable to other diatom- and coccolithophore-dominated new production systems is presently unknown. Assuming the observed response can be extrapolated to new production systems in the ocean, we calculate an excess CO₂ sequestration potential by the biological carbon pump of 116 Pg C until 2100 (for details see Methods). The efficiency of this negative feedback mechanism develops only gradually with rising

CO₂, reaching its maximum strength long after atmospheric CO₂ has transgressed to levels tolerable with regard to climate change and ocean acidification.

Aside from its effect on oceanic carbon sequestration, a more efficient biological carbon pump would lead to an expansion of deep ocean oxygen minimum zones with possible consequences for marine biogeochemical cycling. Increasing C:N ratios would also lower the nutritional value of primary-produced organic matter, which may affect the efficiency of bacterial degradation and zooplankton reproduction²⁵, thus having further implications for marine ecosystem dynamics. Changing carbon-to-nutrient ratios in oceanic primary and export production may thereby act as a forceful driver for ecosystem processes and biogeochemical cycles in the future ocean.

METHODS SUMMARY

The study was conducted between 15 May and 9 June 2005 at the Espesgrend Marine Biological Station (at Raunefjorden, 60.3° N, 5.2° E) of the University of Bergen, Norway. Nine polyethylene mesocosm enclosures (~27 m³, 9.5 m water depth) were moored to a raft, and filled with unfiltered, nutrient-poor, post-bloom fjord water pumped from 13.5 m depth adjacent to the raft. The enclosures were covered by gas-tight tents made of ethylene tetrafluoroethylene foil, which allowed for 95% light transmission of the complete spectrum of sunlight.

The carbonate system in the mesocosms was manipulated by CO₂ aeration to obtain triplicates of three concentrations, 350 µatm (1 × CO₂), 700 µatm (2 × CO₂) and 1,050 µatm (3 × CO₂) (for details see ref. 19). CO₂ aeration of the water column was ended after 3 days, when target CO₂ levels were reached. Continuous flushing of the tents with air adjusted at target CO₂ concentrations ensured that starting values were maintained in the overlying air throughout the experiment.

The water column was stratified by addition and subsequent mixing of 800 litres of freshwater into the upper 5.5 m of the mesocosms, resulting in a salinity (S) gradient of 1.5 p.s.u. between the surface mixed layer (S = 30.6 p.s.u.) and the underlying water column. Continuous mixing of this upper layer by peristaltic pumps (flow rate 450 l h⁻¹) maintained a homogenous distribution of dissolved compounds. To promote the development of a phytoplankton bloom, nitrate and phosphate were added to yield initial concentrations of 14 µmol l⁻¹ NO₃⁻ and 0.7 µmol l⁻¹ PO₄³⁻. Owing to left-over silicate in post-bloom waters, the experiment started at a Si(OH)₄ concentration of 3.2 µmol l⁻¹. After nutrient addition, the development and decline of a phytoplankton bloom was closely monitored over a 24-day period (for details on sampling and measurements see Methods).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 June; accepted 17 September 2007.

Published online 11 November 2007.

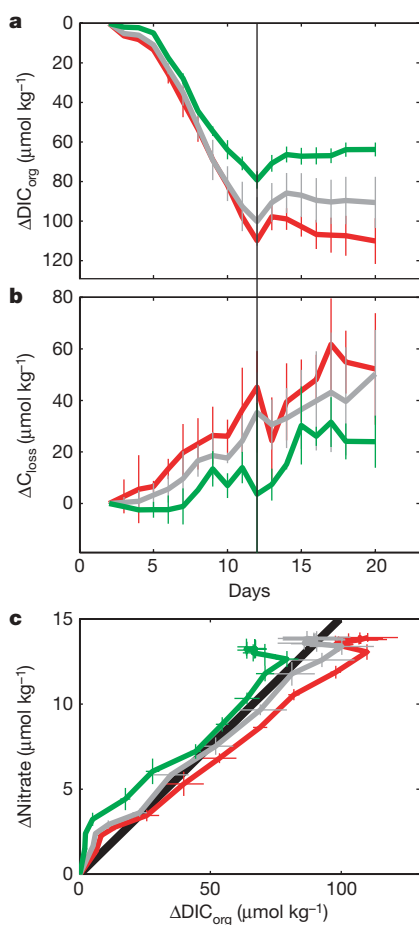


Figure 3 | CO₂ sensitivity of carbon consumption and carbon loss. **a**, Drawdown of dissolved inorganic carbon owing to organic matter production (ΔDIC_{org}) in the upper 5.5 m mixed layer (UML) with time. **b**, Loss of organically bound carbon (ΔC_{loss}) from the UML with time. The vertical line in **a** and **b** indicates time of deep water injection into UML owing to vertical mixing (day 12). **c**, ΔDIC_{org} versus nitrate drawdown in the UML. The black line represents Redfield C:N ratio of 6.6. Values are means of triplicate CO₂ treatments. Error bars denote ± 1 standard deviation; colour code is as in Fig. 1.

1. Sabine, C. L. *et al.* The oceanic sink for anthropogenic CO₂. *Science* **305**, 367–371 (2004).
2. Feely, R. A. *et al.* Impact of anthropogenic CO₂ on the CaCO₃ system in the oceans. *Science* **305**, 362–366 (2004).
3. Caldeira, K. & Wickett, M. E. Anthropogenic carbon and ocean pH. *Nature* **425**, 365 (2003).
4. Langdon, C. *et al.* Effect of elevated CO₂ on the community metabolism of an experimental coral reef. *Glob. Biogeochem. Cycles* **17**, 1011 (2003).
5. Gattuso, J.-P., Frankignoulle, M., Bourge, I., Romaine, S. & Buddemeier, R. W. Effect of calcium carbonate saturation of seawater on coral calcification. *Glob. Planet. Change* **18**, 37–46 (1998).
6. Riebesell, U. & Zondervan, I. Rost, B. Tortell, P. D., Zeebe, R. E. & Morel, F. M. M. Reduced calcification in marine plankton in response to increased atmospheric CO₂. *Nature* **407**, 634–637 (2000).
7. Redfield, A. C., Ketchum, B. H. & Richards, F. A. in *The Sea* 2nd edn (ed. Hill, M. N.) 26–77 (Wiley, New York, 1963).
8. Volk, T. & Hoffert, M. I. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sunquist, E. T. & Broecker, W. S.) Monograph Vol. 32 99–110 (Am. Geophys. Union, Washington DC, 1985).
9. Gruber, N. & Sarmiento, J. L. in *The Sea: Biological-Physical Interactions in the Oceans* Vol. 12 (eds Robinson, A. R., McCarthy, J. J. & Rothschild, B. J.) 337–399 (Wiley, New York, 2002).
10. Houghton, J. T., *et al.* *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, Cambridge, UK, 2001).
11. Raven, J. *et al.* Ocean acidification due to increasing atmospheric carbon dioxide. Policy Document 12/05. *Roy. Soc. Rep.* 12 (2005).
12. Antia, N. J., McAllister, C. D., Parsons, T. R., Stephens, K. & Strickland, J. D. H. Further measurements of primary production using a large-volume plastic sphere. *Limnol. Oceanogr.* **8**, 166–183 (1963).

13. Sambrotto, R. N. *et al.* Elevated consumption of carbon relative to nitrogen in the surface ocean. *Nature* **363**, 248–250 (1993).
 14. Banse, K. Uptake of inorganic carbon and nitrate by marine plankton and the Redfield ratio. *Glob. Biogeochem. Cycles* **8**, 81–84 (1994).
 15. Engel, A., Thoms, S., Riebesell, U., Rochelle-Newall, E. & Zondervan, I. Polysaccharide aggregation as a potential sink of marine dissolved organic carbon. *Nature* **428**, 929–932 (2004).
 16. Engel, A. Direct relationship between CO₂ uptake and transparent exopolymer particles production in natural phytoplankton. *J. Plankton Res.* **24**, 49–53 (2002).
 17. Heemann, C. *Phytoplanktonexsudation in Abhängigkeit von der Meerwasserkarbonatchemie*. Thesis, Univ. Bremen (2002).
 18. Hein, M. & Sand-Jensen, K. CO₂ increases oceanic primary production. *Nature* **388**, 526–527 (1997).
 19. Engel, A. *et al.* Testing the direct effect of CO₂ concentration on a bloom of the coccolithophorid *Emiliania huxleyi* in mesocosm experiments. *Limnol. Oceanogr.* **50**, 493–504 (2005).
 20. Woods, H. A. *et al.* Temperature and the chemical composition of poikilothermic organisms. *Funct. Ecol.* **17**, 237–245 (2003).
 21. Diehl, S., Berger, S. & Wöhrh, R. Flexible algal nutrient stoichiometry mediates environmental influences on phytoplankton and its abiotic resources. *Ecology* **6**, 2931–2945 (2005).
 22. Broecker, W. S. Ocean geochemistry during glacial time. *Geochim. Cosmochim. Acta* **46**, 1689–1705 (1982).
 23. Toggweiler, J. R. Carbon overconsumption. *Nature* **363**, 210–211 (1993).
 24. Körtzinger, A., Koeve, W., Kähler, P. & Mintrop, L. C:N ratios in the mixed layer during the productive season in the Northeast Atlantic Ocean. *Deep-sea Res. I* **48**, 661–688 (2001).
 25. Sterner, R. W. & Elser, J. J. (eds) *Ecological Stoichiometry* (Princeton Univ. Press, Princeton, 2002).
- Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.
- Acknowledgements** We thank the participants of the Pelagic Ecosystem CO₂ Enrichment study (PeECE III, <http://peece.ifm-geomar.de>). We acknowledge J. Egge, J. Nejstgaard and the staff of the Espegrend Marine Biological Station for helping to organize and set up the mesocosms. This work was supported by the EU projects CARBOOCEAN 'Marine carbon sources and sinks assessment' (GOCE), CABANERA and University of Bergen (LOCUS) funding.
- Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to U.R. (uriebesell@ifm-geomar.de).

METHODS

Sampling. Depth-integrated water samples were taken daily at 10:00 by means of a 5-m long, 6-cm diameter tube that was lowered into the mesocosms, closed at the top, pulled up onto the raft and emptied into sampling bottles. Vertical profiles of temperature and salinity inside the mesocosms were obtained using a hand-operated CTD (SAIV A/S, model SD204).

Measurements. CO_2 partial pressure in seawater and the overlying air was measured using an infrared gas analyser (Li-Cor 6262) coupled to an equilibrator²⁶. The p_{CO_2} system was mounted on a cart that was moved between the mesocosms. Water was pumped from the mesocosm at 2 litre min^{-1} and returned to the mesocosm after passing through the equilibrator. Temperature at the inlet of the pump and in the equilibrator was measured simultaneously using platinum RTD digital thermometers. The system was calibrated before and after water analyses with air standards with nominal mixing ratios of 345, 415 and 1,100 p.p.m. Immediately after each measurement of p_{CO_2} , samples for total alkalinity and DIC were taken using the same water pump to fill bottles with ground glass stoppers. Total alkalinity and DIC samples were poisoned with HgCl_2 on collection and were filtered through glass fibre (GF/F) filters before analysis. Total alkalinity was measured using the classical Gran electrotitration method²⁷. The reproducibility of measurements was usually within 4 $\mu\text{mol kg}^{-1}$. DIC was measured by colometric titration²⁸ with a precision of 2 $\mu\text{mol kg}^{-1}$.

For the analysis of phytoplankton pigments, 250–500 ml of the water samples were filtered through 25 mm Whatman GF/F filters. Filters were frozen at -20°C until analysis. For pigment extraction, filters were homogenized in plastic vials (11 ml) together with 1 ml acetone (100%) and a mixture of glass beads (2 and 4 mm) by shaking (5 min) in a cooled Vibrogen cell mill. The extracts were centrifuged ($\sim 850\text{ g}/5,000\text{ r.p.m.}$ for 10 min, cooled at -10°C). The extraction process was performed under dimmed light to prevent photo-oxidation of the pigments. Concentrations of pigments (chlorophyll and carotenoids) were determined by reverse-phase high-performance liquid chromatography²⁹. Identification of pigments was carried out by comparing their retention times and absorption spectra obtained with a diode array spectrophotometer (WATERS) with those of pigment standards. Calibration was carried out with commercially available standards. Calculation of the composition of the phytoplankton communities was executed using the CHEMTAX program³⁰, converting the concentrations of marker pigments to equivalents of chlorophyll *a* with suitable pigment-to-chlorophyll *a* ratios.

Nitrate, nitrite, phosphate and silicate were determined from GF/F-filtered samples with an autoanalyser (AA II)³¹. Ammonium was measured according to ref. 32. POC and PON were measured on 0.25 litre or 0.5 litre samples filtered gently (200 mbar) through precombusted (450°C , 5 h) glass fibre filters (GF/F, Whatman). Filters were fumed overnight with saturated HCl to remove all particulate inorganic carbon, dried for 6 h at 60°C and measured on an elemental analyser (EuroEA 3000, EuroVector). DOC was measured using the high-temperature catalytic oxidation method³³.

Calculations. Inorganic carbon drawdown owing to organic matter production was calculated from measured changes in DIC and alkalinity according to: $\Delta\text{DIC}_{\text{org}} = |\Delta\text{DIC}| - 0.5|\Delta\text{TA}|$, where TA is total alkalinity. ΔDIC was corrected for CO_2 air–sea gas exchange following ref. 34 with chemical enhancement factors given in ref. 35. ΔTA was calculated from changes in DIC and p_{CO_2} using stoichiometric dissociation constants given in ref. 36. Calcium carbonate precipitation was estimated from alkalinity drawdown as $\Delta\text{DIC}_{\text{calc}} = 0.5(\Delta\text{TA} + \Delta\text{nitrate})$. Owing to strong wind and wave action in the fjord on day 12, some water in the mesocosms from below the halocline was mixed into the surface layer. Because surface water was lower in salinity, p_{CO_2} , DIC and total alkalinity owing to the initial freshwater addition, injection of deep water led to a slight increase in p_{CO_2} (Fig. 1a), DIC and total alkalinity on this day. This also affected the calculations of $\Delta\text{DIC}_{\text{org}}$, $\Delta\text{C}_{\text{loss}}$ (Fig. 3a, b) and $\Delta\text{DIC}_{\text{calc}}$ (Supplementary Fig. 3).

Biogenic calcification. Inorganic carbon drawdown owing to biogenic calcification ($\Delta\text{DIC}_{\text{calc}}$) calculated from changes in surface layer alkalinity showed no distinct CO_2 treatment difference during bloom development (Supplementary Fig. 3). Although calcification continued for some days after the peak of the

bloom, this was not depicted in the alkalinity measurements owing to mixing in of higher alkalinity deep water on day 12. Because mixing across the halocline was similar in all CO_2 treatments, the divergence of estimated $\Delta\text{DIC}_{\text{calc}}$ however, indicates a CO_2 effect on calcification after the peak of the bloom. This phase of an *E. huxleyi* bloom, in which nutrient limitation impedes further cell division, is often characterized by excess calcification, leading to shedding of calcite platelets (coccoliths) and their accumulation in the water column. Overall, *E. huxleyi* cell numbers remained much below those typically observed in blooms of this species, which explains the comparatively low amount of calcium carbonate precipitation.

Excess carbon sequestration. Provided that the mechanisms underlying the observed CO_2 sensitivity of carbon consumption and carbon-to-nutrient stoichiometry can be generalized to other diatom- and coccolithophore-dominated new production systems, the excess CO_2 sequestration potential by the ocean's biological carbon pump was calculated for the periods from 1750 to present, and from 1750 to 2100 assuming: business-as-usual CO_2 emissions (IS92a¹⁰) until the end of this century; present-day oceanic carbon export of 12 Pg C yr^{-1} (current estimates range between 8 Pg C yr^{-1} and 16 Pg C yr^{-1})³⁷; and an e-folding time of 100 yr for CO_2 sequestered by biological carbon export to get back into contact with the atmosphere³⁸.

Supplementary Fig. 4a depicts $\Delta\text{DIC}_{\text{org}}$ as a function of atmospheric p_{CO_2} , with the present day value ($\Delta\text{DIC}_{\text{org}}$ at 380 μatm) set to 1. Extrapolating the observed trend to pre-industrial CO_2 levels ($p_{\text{CO}_2} = 280\text{ }\mu\text{atm}$) yields a value of ~ 0.95 ; that is, for a given amount of inorganic nutrients, biological carbon consumption under pre-industrial conditions was about 95% of today's level. On the basis of this relationship, an increase in atmospheric CO_2 from 280 μatm to 380 μatm (present day) corresponds to an excess carbon sequestration of 22 Pg (range 14–29 Pg) for the past 150 yr (Supplementary Fig. 4b). Without this negative feedback mechanism, atmospheric CO_2 would be approximately 11 μatm higher than its present value. By the end of this century, this process would sequester an additional 94 Pg C to the deep ocean, bringing total excess carbon sequestration to 116 Pg C (range 76–154 Pg C). This reduces the increase in atmospheric CO_2 by a total of 58 μatm at 2100.

26. Wanninkhof, R. & Thoning, K. Measurement of fugacity of CO_2 in surface water using continuous and discrete sampling methods. *Mar. Chem.* **44**, 189–205 (1993).
27. Gran, G. Determination of the equivalence point in potentiometric titrations of seawater with hydrochloric acid. *Oceanol. Acta* **5**, 209–218 (1952).
28. Johnson, K. M., Williams, P. J., Brandstrom, L. & Sieburth, J. McN. Colometric total carbon analysis for marine studies: automation and calibration. *Mar. Chem.* **21**, 117–133 (1987).
29. Barlow, R. G., Cummings, D. G. & Gibb, S. W. Improved resolution of mono- and divinyl chlorophylls *a* and *b* and zeaxanthin and lutein in phytoplankton extracts using reverse phase C-8 HPLC. *Mar. Ecol. Prog. Ser.* **161**, 303–307 (1997).
30. Mackey, M. D., Mackey, D. J., Higgins, H. W. & Wright, S. W. CHEMTAX — a program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton. *Mar. Ecol. Prog. Ser.* **144**, 265–283 (1996).
31. Hansen, H. P. & Koroleff, F. in *Methods of seawater analysis* 3rd edn (eds Grasshoff, K., Kremling, K. & Ehrhardt, M.) 159–228 (Wiley VCH, Weinheim, 1999).
32. Holmes, R. M., Aminot, A., K  rouel, R., Hooker, B. A. & Peterson, B. J. A simple and precise method for measuring ammonium in marine and freshwater ecosystems. *Can. J. Fish. Aquat. Sci.* **56**, 1801–1808 (1999).
33. Qian, J. & Mopper, K. Automated high-performance, high-temperature combustion total organic carbon analyser. *Anal. Chem.* **68**, 3090–3097 (1996).
34. Delille, B. et al. Response of primary production and calcification to changes of p_{CO_2} during experimental blooms of the coccolithophorid *Emiliania huxleyi*. *Glob. Biogeochem. Cycles* **19**, GB2023 (2005).
35. Kuss, J. & Schneider, B. Chemical enhancement of the CO_2 gas exchange at a smooth seawater surface. *Mar. Chem.* **91**, 165–174 (2004).
36. Zeebe, R. E. & Wolf-Gladrow, D. CO_2 in seawater: equilibrium, kinetics, isotopes. *Elsevier Oceanogr. Ser.*, **65**, (Elsevier, Amsterdam, 2001).
37. Oschlies, A. Model-derived estimates of new production: new results point towards lower values. *Deep-Sea Res.* **48**, 2173–2197 (2001).
38. Maier-Reimer, E., Mikolajewicz, V. & Winguth, A. Future ocean uptake of CO_2 : interaction between ocean circulation and biology. *Clim. Dynam.* **12**, 711–721 (1996).

Phase-contrast X-ray microtomography links Cretaceous seeds with Gnetales and Bennettitales

Else Marie Friis¹, Peter R. Crane², Kaj Raunsgaard Pedersen³, Stefan Bengtson¹, Philip C. J. Donoghue⁴, Guido W. Grimm⁵ & Marco Stampanoni⁶

Over the past 25 years the discovery and study of Cretaceous plant mesofossils has yielded diverse and exquisitely preserved fossil flowers that have revolutionized our knowledge of early angiosperms¹, but remains of other seed plants in the same mesofossil assemblages^{2,3} have so far received little attention. These fossils, typically only a few millimetres long, have often been charred in natural fires and preserve both three-dimensional morphology and cellular detail. Here we use phase-contrast-enhanced synchrotron-radiation X-ray tomographic microscopy to clarify the structure of small charcoalified gymnosperm seeds from the Early Cretaceous of Portugal and North America. The new information links these seeds to Gnetales (including *Erdtmanithecales*, a putatively closely related fossil group²), and to Bennettitales—important extinct Mesozoic seed plants with cycad-like leaves and flower-like reproductive structures. The results suggest that the distinctive seed architecture of Gnetales, *Erdtmanithecales* and Bennettitales defines a clade containing these taxa. This has significant consequences for hypotheses of seed plant phylogeny by providing support for key elements of the controversial anthophyte hypothesis, which links angiosperms, Bennettitales and Gnetales.

Relationships among extant seed plants (cycads, *Ginkgo*, conifers, Gnetales and angiosperms) remain controversial^{4,5}, but because these are only a small sample of the total diversity in the seed plant clade, new information on extinct seed plants is likely to be crucial to resolve the current impasse^{4–6}. Here we describe new material from a diverse collection of gymnospermous seeds that all have a similar basic structure. These seeds have excellent cellular preservation and are widespread in the Early Cretaceous mesofossil assemblages that we have examined from Portugal and eastern North America. They have been studied previously with standard techniques, but synchrotron radiation X-ray tomographic microscopy, which has been applied successfully to study various microfossils without the need for destructive analysis^{7–9}, provides a more complete understanding of their complex internal structure. For our material, attenuation-based synchrotron-radiation X-ray tomography (SRXTM) does not provide sufficient contrast because of the low absorption of the charcoalified cell walls, whereas phase-contrast X-ray tomographic microscopy (PCXTM, see Methods) provides high resolution at the cellular level.

The fossil seeds are about 0.5–1.8-mm long, obovate to obtriangular or elliptical in longitudinal outline, and distinctly four-angled in transverse section (Figs 1, 2). At the apex there is a central projection, which in some seeds is surrounded by four, pointed, tepal-like structures that project upwards from the four angles. The seeds have a broad stalk, but in most specimens it is broken off. The fossils superficially resemble angiosperm flowers¹⁰, but have also been compared

with seeds of Gnetales and *Erdtmanithecales*^{3,10}. PCXTM establishes that the fossils are small seeds comprised of two distinctly different layers surrounding the nucellus: an inner thin, membranous, integument, formed by thin-walled cells, and a robust outer, sclerenchymatous, seed envelope that completely encloses the integument except for the micropylar opening.

The integument is slightly shrunken in all specimens examined but there is no well-developed nucellus membrane or megaspore membrane. The integument and outer envelope are free for their full length and are attached only by a broad connection at the base (Fig. 1, left panel). The integument is free from the nucellus in its upper part and tapers into a long, narrow micropylar tube that is circular in transverse section. The micropylar canal is open towards the apex, and the integument in this region consists of one or two layers of small cells (Fig. 1, CS1). In the middle part and towards the base, the micropylar canal is closed by expanded tissue of the integument, which, at this level, consists of several cell layers. In particular, the cells of the inner epidermis are expanded radially to form a prominent ring that closes the micropylar canal (Fig. 1, CS2). At the base of the micropyle, the integument opens again, but the inner lining is irregular and the inner cells are broken down (Fig. 1, CS3).

The envelope is mainly composed of a hard sclerenchymatous tissue. At the apex this tissue forms a distinct, central projection around the micropylar tube formed by the integument. It also forms the protrusions that extend from the corners of the seed (Fig. 1a–d). In some seeds these merge proximally to form a ring-like structure (Fig. 1a); in others they are thinner and more discrete. There is considerable diversity in seed size and the form of the outer surface among this group of seeds.

The envelope has an inner layer of narrowly elongated sclerenchyma cells that are transversely aligned around the main body of the seed. Towards the apex these sclerenchyma cells are expanded radially (Figs 1d, 2c). The inner surface of the envelope is smooth and non-papillate, including around the micropyle (Figs 1d, 2c). The outer surface has irregular transverse ridges, which give the seed a wrinkled appearance (Figs 1b, c, 2a). The cells comprising this layer are elongated and aligned vertically. Occasional remains of parenchymatous tissues outside the wrinkled surface of the sclerenchyma layer suggest that the seed envelope had a thin, fleshy outer covering.

Only Gnetales (extant and extinct), *Erdtmanithecales* (extinct) and Bennettitales (extinct) are known to have seeds of this kind with an additional seed envelope and the integument extended into a long, narrow micropylar tube. In all three extant genera of Gnetales, the integument is thin and fused to the nucellus for most of its length. At the base, the fused integument is broadly attached and surrounded by one (*Ephedra*, *Welwitschia*) or two (*Gnetum*) envelopes, by which it is

¹Departments of Palaeobotany and Palaeozoology, Swedish Museum of Natural History, SE-104 05 Stockholm, Sweden. ²Department of the Geophysical Sciences, University of Chicago, Chicago, Illinois 60637, USA. ³Department of Earth Sciences, University of Aarhus, DK-8000 Århus C, Denmark. ⁴Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK. ⁵Institute of Geosciences, Eberhard-Karls-University, D-72076 Tübingen, Germany. ⁶Swiss Light Source, Paul Scherrer Institut, CH-5232 Villigen, Switzerland.

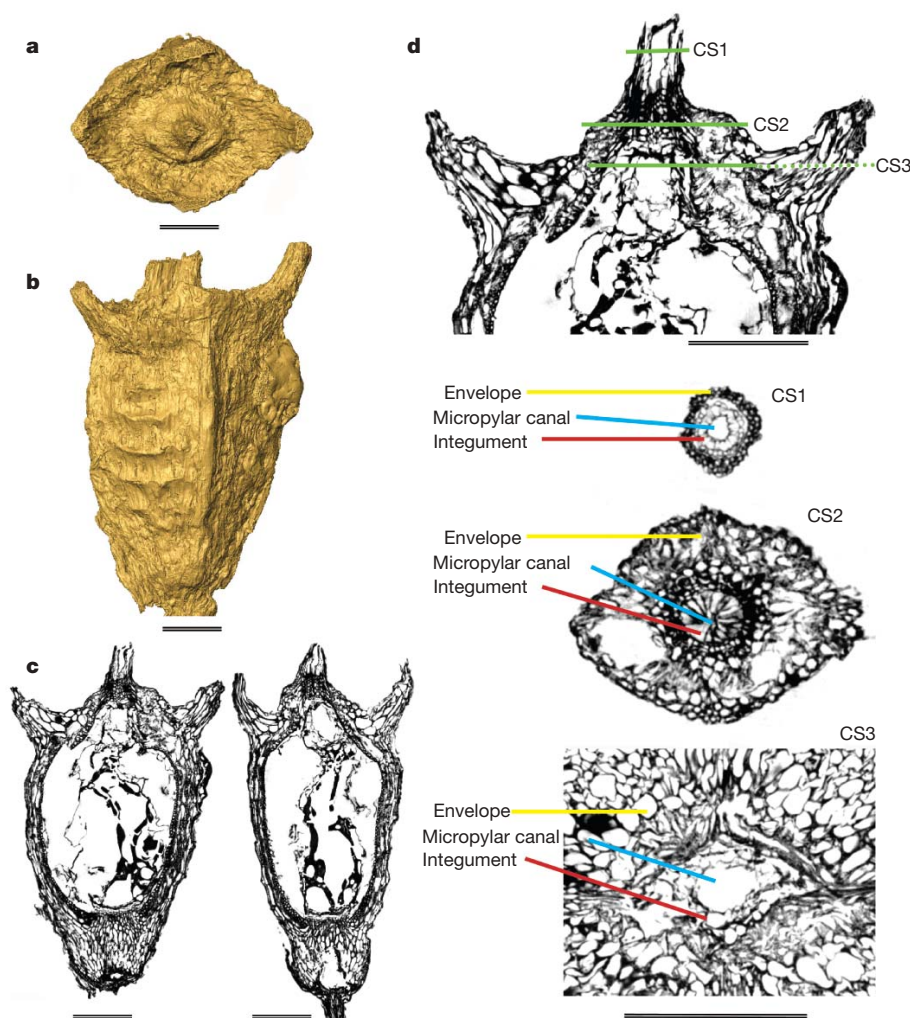


Figure 1 | Early Cretaceous (Barremian–Aptian) seed from Catefica, Portugal, with affinities to Gnetales and Bennettitales. S153152 (see Methods). **a, b**, Tomographic reconstructions of seed morphology in apical (**a**) and lateral (**b**) view showing quadrangular shape, micropylar protrusion and four apical extensions of the envelope. **c, d**, PCXTM images of internal structures. **c**, Longitudinal sections (perpendicular to each other) showing membranous inner tissues (megaspore membrane, nucellus and integument) surrounded by sclerenchymatous tissue of the envelope. **d**, Longitudinal section showing the micropylar area and position of transverse sections (CS1, open micropylar canal; CS2, closed canal; CS3, base of micropylar region). Scale bar in all figures, 200 μ m (CS1, CS2 and CS3 share the same scale bar).

almost entirely enclosed. In *Ephedra*, cells lining the micropylar canal are small and uniform in size and shape. At maturity, the micropylar canal is closed by a hardened mucilaginous secretion¹¹. In *Gnetum*, *Welwitschia* and the fossil seeds described here, the micropylar canal is open distally, but blocked in the middle and also proximally by a multicellular closing tissue formed by the expanded integument^{12–14}.

In *Ephedra* and *Gnetum*, the envelope surrounding the integument has an inner sclerenchymatous and an outer fleshy zone; in *Welwitschia*, this envelope forms a pronounced membranous wing; in extant *Ephedra*, the sclerenchymatous envelope is usually

bisymmetric or sometimes triangular in cross-section. However, several Early Cretaceous *Ephedra* seeds (*E. portugallica* and *E. drewriensis*) have a four-parted envelope that is very similar to that of the fossil seeds described here³. Also similar to the charcoallified seeds, the surface of the sclerenchymatous zone in extant *Ephedra* (for example, *E. rhytidosperra*) and extinct ephedroid seeds is sometimes transversely ribbed^{3,15}. However, seeds of *Ephedra* differ from the material described here in lacking the micropylar closing tissue and in having a distinct papillate inner lining on the outer envelope around the micropylar tube.

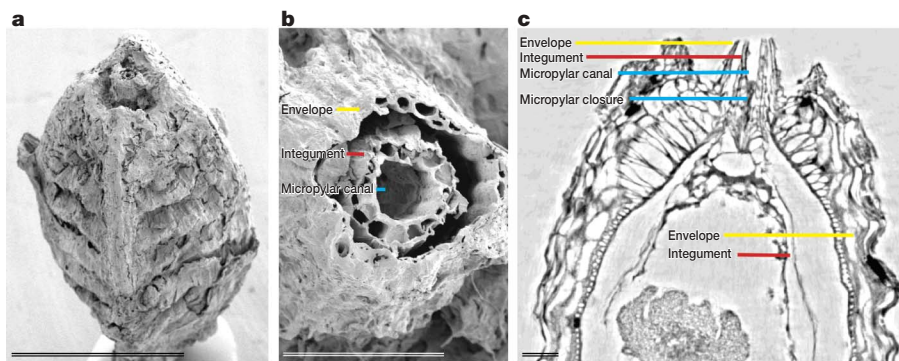


Figure 2 | Early Cretaceous (mid-Albian) seed from the Puddledock locality, USA, with affinities to Gnetales and Bennettitales. PP53361 (see Methods). **a, b**, Scanning electron microscope images of external morphology; oblique lateral view (**a**) showing the ribbed surface of the sclerenchymatous seed envelope and remains of the fleshy outer layer; scale

bar, 500 μ m; apical view (**b**) showing details of the micropylar extension with the seed envelope surrounding the integument and micropylar canal; scale bar, 50 μ m. **c**, PCXTM image (longitudinal) of internal seed structures showing remains of the embryo sac and nucellus surrounded by thin integument and thick sclerenchymatous seed envelope; scale bar, 50 μ m.

Seeds of Erdtmanithecates (*Erdtmanispermum*) from the earliest Cretaceous are very similar to those of *Ephedra*² and have a three-parted sclerenchymatous envelope, which is sometimes faintly ribbed externally. However, they differ from seeds of extant and extinct *Ephedra* by the presence of distinctive *Eucommiidites*-type pollen in the micropyle and by the lack of a papillate lining on the outer envelope around the micropylar tube. Pollen organs of Erdtmanithecates also differ from those of extant Gnetales. Unfortunately, in *Erdtmanispermum* seeds cellular details of the micropyle closure are not clear.

Seeds of Bennettitales are also small except for several Triassic forms^{16,17}. In all cases the seeds are borne on stalks in conical to globose heads, between densely crowded interseminal scales. Dispersed seeds have never been recognized and, despite the excellent preservation of many petrified specimens, the organization of the seeds remains controversial. Bennettitalean seeds have been interpreted as having a single integument and no other additional coverings^{18,19}, or as having a thin integument surrounded by an additional envelope^{16,17,20}. Interseminal scales have been interpreted as aborted ovules, and the entire ovulate structure has been considered to consist of naked seeds (fully developed and aborted) with no supporting structures^{16,17,20}.

Comparison of the charcoalfied seeds from Portugal and North America with petrified seeds of Bennettitales from ovulate heads^{18,19,21} reveals striking similarities. As in the charcoalfied fossils, there is an outer layer with a sclerenchymatous and parenchymatous zone. In some species the sclerenchymatous layer is distinctly four- to five-angled in transverse section, in others it is five- or sometimes six-angled. The outer surface of the sclerenchymatous layer is irregularly ribbed in several taxa. Also, in *Cycadeoidea*, the sclerenchymatous layer often extends apically from the corners of the seed to form extensions that flank the micropylar area, as in the charcoalfied fossil seeds (see Supplementary Information).

Petrified seeds of *Cycadeoidea morierei* show especially well-preserved cellular details that can be compared directly with the charcoalfied fossils and seeds of Gnetales^{12,13}. It is clear that the sclerenchyma layer in *C. morierei* surrounds a discrete micropylar tube with an inner epidermis of prominent radially expanded cells, which forms a ring around the open micropylar canal²¹. Towards the middle and base of the micropylar canal, closure is by a multicellular tissue, and at the base of the canal, disintegration of the central cells forms an irregular cavity²¹. The proliferated cells in the middle of the micropyle are identical to the closing tissue observed in *Gnetum* and *Welwitschia*¹². The details revealed by PCXTM show that this layer is not a nucellar plug or part of the nucellus^{18,19}. In *Cycadeoidea morierei*, there is a thin parenchymatous membrane inside the sclerenchymatous layer, which is similar to the integument seen in the charcoalfied seeds and seeds of Gnetales.

On the basis of anatomical and positional similarities, we equate the sclerenchymatous layer in petrified seeds of Bennettitales to the very similar sclerenchymatous envelope that surrounds the integument in our charcoalfied seeds. This outer envelope is also visible in some Bennettitales, preserved as compressions^{16,17,20}, *Erdtmanispermum*,

Ephedra and *Welwitschia*, and *Gnetum* (middle layer). In Gnetales, this envelope is commonly interpreted as a pair of modified bracts²⁰. The seed envelope in extinct Gnetales, as well as in *Erdtmanispermum*, Bennettitales and the charcoalfied seeds, is also interpreted most straightforwardly as two or more bracts. This has profound implications for the morphological interpretation of bennettitalean 'flowers'—it suggests that the ovulate 'flowers' are an aggregation of strongly condensed bracteate reproductive axes, which opens up new possibilities for clarifying homologies with structures in other seed plants.

The seed architecture elucidated here for the charcoalfied seeds, which also occurs in extant and fossil Gnetales, Erdtmanithecates and Bennettitales, can be summarized as: seeds orthotropous, with a single integument extended into a long micropylar tube; micropylar canal closed after fertilization by expanded integumentary tissue (or in *Ephedra* by mucilaginous secretion); integument thin, weakly developed; nucellus and megaspore membrane poorly developed or lacking; and integument surrounded by a robust outer envelope comprised in part of sclerenchymatous tissues. The possibility that this seed architecture comprises a synapomorphy, or group of synapomorphies, for a bennettitalean–erdtmanithecatean–gnetalean clade (BEG clade) is also supported by the occurrence of pollen grains with a granular infratectal layer in the pollen wall in all three taxa, and the presence of paracytic stomata in Bennettitales, Erdtmanithecates and *Welwitschia*–*Gnetum*. In the context of current phylogenetic analyses of extant and fossil seed plants, these characters provide further support for some elements of the anthophyte grouping recognized by previous studies^{20,22}, and most recently in refs 4 and 5 (see also Supplementary Information).

The new evidence presented here links the charcoalfied seeds to Gnetales, Erdtmanithecates and Bennettitales, and supports recognition of the BEG clade (Fig. 3). It emphasizes the great diversity of the group from which extant *Ephedra*, *Welwitschia* and *Gnetum* emerged, and the extent to which they are relictual. It also supports earlier ideas of a close relationship between Bennettitales and Gnetales^{11,20,23}, but in previously unsuspected ways. Testing this hypothesis will need renewed efforts to resolve seed plant phylogeny and to reconcile morphological results with evidence from molecular phylogenetics, which currently conflicts with the anthophyte grouping²⁴. Such efforts will also require exploration of the impacts of the trade-off between increased taxon versus increased character sampling on phylogenetic pattern and resolution. Our results also raise the question of how the close similarity among the ovulate structures of Gnetales, Erdtmanithecates and Bennettitales relates to the situation in other seed plants, especially in angiosperms and other possible anthophytes. A next generation of phylogenetic analyses will need to incorporate a greater range of fossil taxa, as well as new kinds of material and character systems, but the consistent occurrence of two layers around the nucellus in the ovules of all three BEG lineages is similar to the typical bitegmic ovules of angiosperms, and is worthy of closer investigation.

METHODS SUMMARY

The fossil seeds were collected at several localities in the Western Portuguese Basin and the Potomac Group of eastern North America, which range in age from late Barremian/early Aptian to mid-Albian. The fossils were extracted from the sediments by sieving in water, and cleaned using established techniques. The seeds were imaged on the X-ray microscope at the TOMCAT beamline of the Swiss Light Source of the Paul Scherrer Institute. We used phase-contrast X-ray tomographic microscopy (PCXTM) to enhance the quality of the tomograms^{25,26}. The enhanced contrast of this X-ray imaging technique arises from the phase shift induced in the X-ray wave as it passes through the sample. This method differs from other approaches^{27,28} because it couples high resolution and high sensitivity with speed: phase information is obtained by a single tomographic scan. The underlying procedure is based on the assumption of low and homogeneous absorption, but the quality of the resulting images is largely sufficient to perform optimal segmentation and further post-processing. The original algorithm was introduced by Bronnikov²⁹, further refined by Gureyev *et al.*³⁰ and implemented experimentally by Grosio *et al.*²⁵.

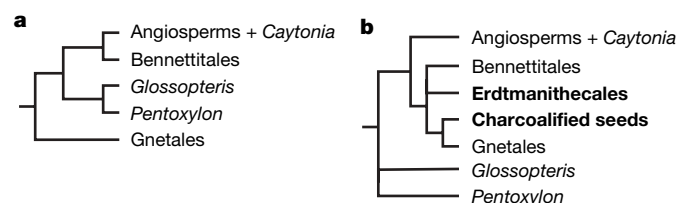


Figure 3 | The phylogenetic relationships among 'anthophytes'.

a, Topology as for ref. 5, with the Gnetales as the sister clade to the 'higher' seed ferns (*Caytonia*, *Glossopteris* and *Pentoxylon*), Bennettitales and angiosperms. **b**, Topology, under inclusion of the charcoalfied seeds and Erdtmanithecates, links the Bennettitales with the Gnetales. For details see Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 July; accepted 19 September 2007.

- Friis, E. M., Pedersen, K. R. & Crane, P. R. Cretaceous angiosperm flowers: Innovation and evolution in plant reproduction. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **232**, 251–293 (2006).
- Pedersen, K. R., Crane, P. R. & Friis, E. M. Pollen organs and seeds with *Eucommiidites* pollen. *Grana* **28**, 279–294 (1989).
- Rydin, C., Pedersen, K. R., Crane, P. R. & Friis, E. M. Former diversity of *Ephedra* (Gnetales): evidence from Early Cretaceous seeds from Portugal and North America. *Ann. Bot. (Lond.)* **98**, 123–140 (2006).
- Doyle, J. A. Seed ferns and the origin of angiosperms. *J. Torrey Bot. Soc.* **133**, 169–209 (2006).
- Hilton, J. & Bateman, R. M. Pteridosperms are the backbone of seed-plant phylogeny. *J. Torrey Bot. Soc.* **133**, 119–168 (2006).
- Bateman, R. M., Hilton, J. & Rudall, P. J. Morphological and molecular phylogenetic context of the angiosperms: contrasting the 'top-down' and 'bottom-up' approaches used to infer the likely characteristics of the first flowers. *J. Exp. Bot.* **57**, 3471–3503 (2006).
- Donoghue, P. C. J. et al. Synchrotron X-ray tomographic microscopy of fossil embryos. *Nature* **442**, 680–683 (2006).
- Tafforeau, P. et al. Applications of X-ray synchrotron microtomography for non-destructive 3D studies of paleontological specimens. *Appl. Phys. A* **83**, 195–202 (2006).
- Tafforeau, P., Bentele, I., Jaeger, J. J. & Martin, C. Nature of laminations and mineralization in rhinoceros enamel using histology and X-ray synchrotron microtomography: Potential implications for palaeoenvironmental isotopic studies. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **246**, 206–227 (2007).
- Friis, E. M., Pedersen, K. R. & Crane, P. R. Angiosperm floral structures from the Early Cretaceous of Portugal. *Plant Syst. Evol.* **8** (suppl.), 31–49 (1994).
- Thoday, M. G. The female inflorescence and ovules of *Gnetum africanum*, with notes on *Gnetum scandens*. *Ann. Bot. (Lond.)* **25**, 1101–1135 (1911).
- Berridge, E. M. On some points of resemblance between Gnetalean and Bennettitean seeds. *New Phytol.* **10**, 140–144 (1911).
- Pearson, H. H. W. *Gnetales* (Cambridge Univ. Press, Cambridge, 1929).
- Martens, P. *Handbuch der Pflanzenanatomie*, vol. 12, pt 2: *Les gnétophytes* (eds Zimmermann, W., Carlquist, S., Ozenda, P. & Wulff, H. D.) (Gebrüder Borntraeger, Berlin, Stuttgart, 1971).
- Yang, Y., Geng, B.-Y., Dilcher, D. L., Chen, Z.-D. & Lott, T. A. Morphology and affinities of an Early Cretaceous *Ephedra* (Ephedraceae) from China. *Am. J. Bot.* **92**, 231–241 (2005).
- Harris, T. M. The fossil flora of Scoresby Sound East Greenland. Part 3: Caytoniales and Bennettitales. *Medd. Grøn.* **85**, 1–133 (1932).
- Pedersen, K. R., Crane, P. R. & Friis, E. M. Morphology and phylogenetic significance of *Vardekloeftia* Harris (Bennettitales). *Rev. Palaeobot. Palynol.* **60**, 7–24 (1989).
- Stokes, M. C. New Bennettitean cones from the British Cretaceous. *Phil. Trans. R. Soc. B* **208**, 389–440 (1918).
- Rothwell, G. W. & Stockey, R. A. Anatomically preserved *Cycadeoidea* (Cycadeoidaceae), with a reevaluation of systematic characters for the seed cones of Bennettitales. *Am. J. Bot.* **89**, 1447–1458 (2002).
- Crane, P. R. Phylogenetic analysis of seed plants and the origin of angiosperms. *Ann. Mo. Bot. Gard.* **72**, 716–793 (1985).
- Lignier, O. *Végétaux fossiles de Normandie. Structure et affinités du Bennettites morierei* Sap. & Mar. (sp.) 1–78 (E. Lanier, Caen, 1894).
- Doyle, J. A., Donoghue, M. J. & Zimmer, E. A. Integration of morphological and ribosomal RNA data on the origin of angiosperms. *Ann. Mo. Bot. Gard.* **81**, 419–450 (1994).
- Arber, E. A. N. & Parkin, J. On the origin of angiosperms. *J. Linn. Soc. Bot.* **38**, 29–80 (1907).
- Burleigh, J. G. & Mathews, S. Phylogenetic signal in nucleotide data from seed plants: implications for resolving the seed plant tree of life. *Am. J. Bot.* **91**, 1599–1613 (2004).
- Grosio, A., Abela, R. & Stampanoni, M. Implementation of a fast method for high resolution phase contrast tomography. *Opt. Express* **14**, 8103–8110 (2006).
- Bronnikov, A. V. Reconstruction formulas in phase-contrast tomography. *Opt. Commun.* **171**, 239–244 (1999).
- Cloetens, P. et al. Holotomography: Quantitative phase tomography with micrometer resolution using hard synchrotron radiation X rays. *Appl. Phys. Lett.* **75**, 2912–2914 (1999).
- Weikamp, T. et al. X-ray phase imaging with a grating interferometer. *Opt. Express* **13**, 6296–6304 (2005).
- Bronnikov, A. V. Theory of quantitative phase-contrast computed tomography. *J. Opt. Soc. Am. A* **19**, 472–480 (2002).
- Gureyev, T. E., Paganin, D. M., Myers, G. R., Nesterets, Y. I. & Wilkins, S. W. Phase-and-amplitude computer tomography. *Appl. Phys. Lett.* **89**, 034102–1–034102-3 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F. Marone of the Swiss Light Source for valuable help in setting up the software on the cluster, and T. Hultgren and N. J. Gostling for help with data analyses. We also thank the Natural History Museum, London, for the loan of petrified specimens, J.-P. Rioult for information on the Lignier fossils, and J. Hilton for helpful suggestions. This study was funded by the Swiss Light Source (P.C.J.D. and S.B.), European Union FP6, as well as by grants from the Swedish Research Council (E.M.F.), the Carlsberg Foundations (K.R.P. and E.M.F.) and the National Science Foundation (P.R.C.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.M.F. (else.marie.friis@nrm.se) or M.S. (marco.stampanoni@psi.ch).

METHODS

Geological occurrence and preparation. The fossil seeds were collected from unconsolidated sands and clay at several localities in the Western Portuguese Basin and the Potomac Group of eastern North America. These localities range in age from the late Barremian/early Aptian to the mid-Albian. The fossils were extracted from the sediments by sieving in water. They were then cleaned in HF and HCl, followed by rinsing in water. The two specimens shown here were analysed using PCXTM without further treatment. Other specimens were first studied using scanning electron microscopy (SEM) and then remounted for PCXTM. All specimens for SEM were coated with gold. Nail polish was used as mounting media for both PCXTM and SEM. The fossil material is housed in the palaeobotanical collections of the Swedish Museum of Natural History (S) and the geological collections of the Field Museum, Chicago (PP).

Phase-contrast-enhanced synchrotron-radiation X-ray tomographic microscopy. Standard attenuation-based synchrotron radiation X-ray tomography (SRXTM) has been applied successfully to study various microfossils^{7–9}, but does not provide sufficient contrast in this fossil material because of the extremely low absorption of the charcoallified cell walls. We used phase-contrast X-ray tomographic microscopy (PCXTM) to enhance the quality of the tomograms^{25,26}. The enhanced contrast of this X-ray imaging technique is not dependant on simple attenuation but arises from the phase shift induced in the X-ray wave as it passes through the sample. The method used here differs from other approaches^{27,28} because it couples high resolution and high sensitivity with speed. The method yields the three-dimensional distribution of the phase (refractive index) of a weakly absorbing object from a single tomographic scan, minimizing measurement time and delivered dose.

Even though the underlying numerical approach is based on the assumption of flow and homogeneous absorption, the quality of the resulting images is largely sufficient to perform optimal segmentation and further post-processing. We provide here a very short mathematical description of the physics underlying this method.

Under the aforementioned assumption, the intensity distribution at distance $z = d$ and angle of rotation θ is approximated by the following simplified ‘Transport of Intensity’ equation²⁶:

$$I_{\theta,z=d}(x,y) = I_{\theta,z=0}(x,y) \left[1 - \frac{\lambda d}{2\pi} \nabla^2 \phi_{\theta}(x,y) \right] \quad (1)$$

where $\phi_{\theta}(x,y)$ is the phase function of the object. The goal of the method is to reformulate equation (1) to obtain $\delta(x,y,z)$, representing the real part of the complex index of refraction $n = 1 - \delta + i\beta$, from the knowledge of $I_{\theta,z=d}(x,y)$ for $\theta \in [0, \pi]$.

Expressing equation (1) as $\nabla^2 \phi_{\theta}(x,y) = -\frac{2\pi}{\lambda d} g_{\theta}(x,y)$ with $g_{\theta}(x,y) = \frac{I_{\theta,z=d}(x,y)}{I_{\theta,z=0}(x,y)} - 1$, applying the 3D Radon transform (denoted by the symbol \wedge) and calculating finally the second derivative with respect to the variable $s = x \sin \omega + y \cos \omega$ one gets:

$$\frac{\partial^2}{\partial s^2} \delta(s, \theta, \omega) = -\frac{1}{d} \wedge g_{\theta}(s, \omega) \quad (2)$$

Expression (2) is a theorem that states that from the 2D Radon transform of the measured value g , one can directly find the 3D Radon transform of δ .

Applying the 3D Radon transform and developing the full derivatives, it can be shown that:

$$\delta(x,y,z) = \frac{1}{4\pi^2 d} \int_0^{\pi} (q^{**} g_{\theta}) d\theta \quad (3)$$

where $q(x,y) = \frac{|y|}{x^2 + y^2}$ and $**$ defines the two-dimensional convolution operator.

According to the convolution theorem and taking empirically into account the effects of the, always present, non-zero absorption, equation (3) can be evaluated via Fourier space transformation with the following low-pass filter:

$$q(\xi, \eta) = \frac{|\xi|}{\xi^2 + \eta^2 + \alpha_{\text{exp}}} \quad (4)$$

The values of α_{exp} to be used are found by using a semi-empirical (simulations–experiment) approach which goes beyond the purposes of this article.

We modified Grosio’s implementation²⁵ to automatically estimate the factor α_{exp} for each angular projection, optimizing the correction for every angle and therefore enhancing the final quality of the reconstructed slices. Fast Fourier transforms and filtering are done on-the-fly on a 16 nodes Linux cluster during data acquisition. After tomographic reconstruction, the relative values of δ are available: even though this method does not directly provide an absolute value for δ , it definitely delivers images with increased contrast sufficient to perform segmentation.

Slice data. The seeds were imaged using a 10 \times objective on the X-ray microscope at the TOMCAT beamline of the Swiss Light Source of the Paul Scherrer Institut. Exposure time at 20 keV was 450 ms, and 1,501 projections were acquired equi-angularly over 180°. Projections were post-processed and rearranged into flat- and darkfield-corrected sinograms, and reconstruction was performed on a 32-nodes Linux PC farm. Isotropic voxel dimensions are 0.7 μm . Slice data were analysed and manipulated using AMIRA (www.tgs.com) software for computed tomography.

An antidepressant that extends lifespan in adult *Caenorhabditis elegans*

Michael Petrascheck¹, Xiaolan Ye¹ & Linda B. Buck¹

The mechanisms that determine the lifespan of an organism are still largely a mystery¹. One goal of ageing research is to find drugs that would increase lifespan and vitality when given to an adult animal. To this end, we tested 88,000 chemicals for the ability to extend the lifespan of adult *Caenorhabditis elegans* nematodes. Here we report that a drug used as an antidepressant in humans increases *C. elegans* lifespan. In humans, this drug blocks neural signalling by the neurotransmitter serotonin. In *C. elegans*, the effect of the drug on lifespan is reduced or eradicated by mutations that affect serotonin synthesis, serotonin re-uptake at synapses, or either of two G-protein-coupled receptors: one that recognizes serotonin and the other that detects another neurotransmitter, octopamine. *In vitro* studies show that the drug acts as an antagonist at both receptors. Testing of the drug on dietary-restricted animals or animals with mutations that affect lifespan indicates that its effect on lifespan involves mechanisms associated with lifespan extension by dietary restriction. These studies indicate that lifespan can be extended by blocking certain types of neurotransmission implicated in food sensing in the adult animal, possibly leading to a state of perceived, although not real, starvation.

The short lifespan of *C. elegans* (~3 weeks) can be increased by dietary restriction—an effect seen in many organisms—as well as by alterations in a variety of genes, some with analogous effects in fruitflies and/or mice^{2,3}. Several chemicals have also been found to increase lifespan in invertebrates, including one identified by testing *C. elegans* with 19 compounds and another that also increases lifespan in yeast and fish^{4–8}. However, no large-scale screens have been conducted for longevity enhancing drugs. The identification of such drugs could provide additional insights into ageing mechanisms and ultimately point to drugs suitable for testing in mammals.

To search for compounds that increase lifespan when given to adult *C. elegans*, we tested 88,000 diverse chemicals on animals grown in liquid medium in 384-well plates. Starting at day 1 of adulthood, animals in each well were continuously exposed to a single chemical at 30–90 μ M. On the basis of the fraction of live animals per well relative to controls, 1,083 chemicals were retested on larger populations. Of these, 115 compounds statistically increased lifespan, with 13 increasing lifespan by 30–60%, 18 by 20–29%, 27 by 10–19% and 57 by 3–9%. The number of screened chemicals that actually entered the animal is unknown as is the number that increased lifespan by acting on the same endogenous target(s).

One compound that increased lifespan by 20%, '272N18' (3-(3-nitrophenyl)-11-phenyl-2,3,4,5,10,11-hexahydro-1H-dibenzo[b,e][1,4]diazepin-1-one dihydrochloride), is structurally related to certain antidepressant drugs (Fig. 1a). In humans, these antidepressants affect intercellular signalling by serotonin, a neurotransmitter found in many animals, including *C. elegans*^{9–11}.

We subsequently tested *C. elegans* with 20 different compounds that affect serotonin signalling pathways (Supplementary Table 1)^{9,10}.

Four compounds increased lifespan by 20–33%: mianserin, mirtazapine, methiothepin and cyproheptadine (Fig. 1 and Supplementary Table 1). In humans, all four compounds are antagonists of serotonin 2 (5-HT₂) receptors and, to a variable extent, of certain other biogenic amine receptors⁹ (see also <http://pubchem.ncbi.nlm.nih.gov/>). Mianserin and mirtazapine are used to treat depression and cyproheptadine is used to treat migraine and allergies, whereas

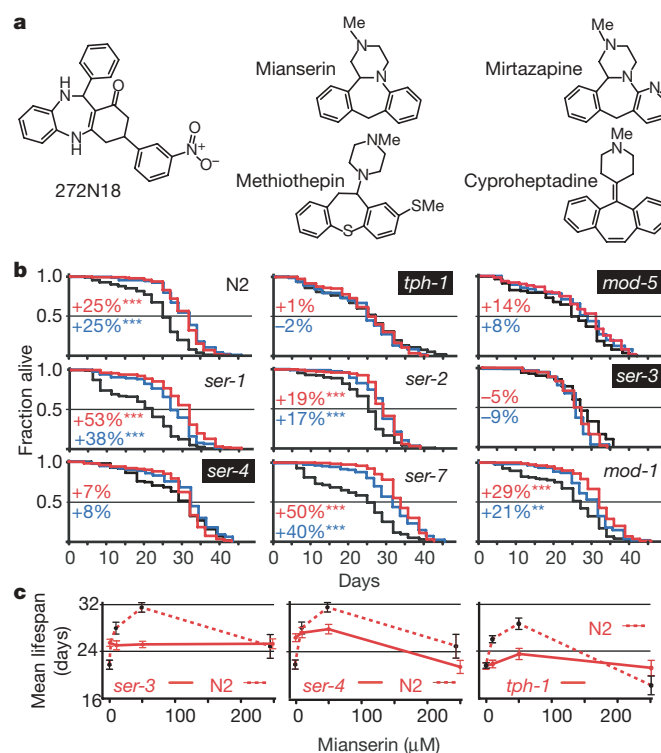


Figure 1 | Increases in *C. elegans* lifespan by human serotonin receptor antagonists require serotonin and octopamine signalling. **a**, Structures of one chemical (272N18) and four serotonin receptor antagonists that increased *C. elegans* lifespan. **b**, Survival curves from representative experiments show the fraction of wild-type (N2) or mutant animals alive at different ages when given 50 μ M mianserin (red), 10 μ M methiothepin (blue), or no drug (black). Genes required for lifespan extension by the drugs are highlighted in black. Percentage changes in lifespan versus untreated controls are indicated for each drug. Asterisks indicate significant increases (triple asterisk, $P \leq 0.0001$; double asterisk, $P \leq 0.001$; no asterisk, $P > 0.01$; sample sizes were ≥ 50). **c**, Mean lifespan as a function of mianserin concentration. Maximum increase in lifespan of N2 animals (dotted red line) was seen at 50 μ M mianserin concentration. The lifespans of *ser-3*, *ser-4* and *tph-1* mutants tested in parallel (solid red lines) were largely unaffected. Error bars indicate s.e.m. See Supplementary Table 2 for details.

¹Howard Hughes Medical Institute, Basic Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109-1024, USA.

methiothepin is not used clinically (ref. 12 and <http://pubchem.ncbi.nlm.nih.gov/>). The effect of mianserin on lifespan was dose dependent with a $31 \pm 3\%$ maximal increase in lifespan obtained with $50 \mu\text{M}$ drug (Fig. 1c). Other antidepressants tested did not increase lifespan, including several serotonin-specific re-uptake inhibitors (Supplementary Table 1).

To determine whether mianserin increases *C. elegans* longevity by altering serotonin signalling, we tested the drug on animals with mutations in different serotonin signalling components. In these and most other studies reported below, we also tested methiothepin or cyproheptadine and obtained results similar to those with mianserin (Supplementary Table 4).

Mianserin failed to increase lifespan in *tph-1(mg280)* mutant nematodes that lack tryptophan hydroxylase, a key enzyme in serotonin synthesis¹³ (Fig. 1b, c and Table 1). Furthermore, in animals lacking the serotonin re-uptake transporter MOD-5, maximum lifespan extension by mianserin was 14% compared to $31 \pm 3\%$ in wild-type animals¹⁴ (Fig. 1b, c and Table 1). Thus, full lifespan extension by mianserin requires both serotonin production and its re-uptake at synapses.

We also tested the drug on animals mutant for each member of the serotonin/octopamine/tyramine receptor family of G-protein-coupled receptors (GPCRs) (SER-1, SER-2, SER-3, SER-4 and SER-7) or for MOD-1, the serotonin-gated chloride channel^{15,16}. Mianserin binds both SER-2 and MOD-1, and genetic studies suggest that it might bind SER-4 (refs 10, 15). The *ser-1(ok345)* mutant is reportedly long lived¹⁷. Mianserin caused increases in lifespan similar to, or greater than, those seen in wild-type animals in *ser-1(ok345)*, *ser-2(pk1357)*, *ser-7(tm1325)* and *mod-1(ok103)* mutants.

In contrast, mianserin did not extend the lifespan of two different *ser-3* mutants (Fig. 1 and Table 1) and it produced only a $7 \pm 2\%$ lifespan increase in the *ser-4(ok512)* mutant. When a wild-type *ser-4* transgene was expressed in the *ser-4(ok512)* mutant, mianserin caused a 16% lifespan increase versus a 4% increase in the *ser-4(ok512)* mutant alone (Supplementary Fig. 1). These results were not due to decreased mianserin uptake by the mutants, as pre-incubating them with the drug decreased serotonin-induced egg laying¹⁰ (see below). SER-4 is activated by serotonin¹⁵, and genetic studies suggest that SER-3 is activated by octopamine, the proposed invertebrate equivalent of noradrenalin^{16,18}. Our results thus indicate that lifespan extension by mianserin requires both a serotonin receptor, SER-4, and a probable octopamine receptor, SER-3^{15,16}.

To examine directly mianserin's effect on SER-3 and SER-4 receptors, we used calcium imaging¹⁹. HEK293 cells were transfected with a SER-3 expression vector²⁰ or with SER-4 and $\text{G}\alpha 15^{21}$ expression

vectors. Cells expressing SER-3 responded to as little as 10 nM octopamine and to tyramine at $10 \mu\text{M}$, but showed no response to serotonin (Fig. 2a). SER-4⁺ cells responded to serotonin at concentrations as low as 100 nM (Fig. 2c). Neither SER-3⁺ nor SER-4⁺ cells responded to mianserin.

We next asked whether mianserin acts as a SER-3 or SER-4 antagonist. Receptor-expressing HEK293 cells were pre-incubated with mianserin (or methiothepin) and then exposed to octopamine (SER-3) or serotonin (SER-4) in the presence of the drug. Mianserin and methiothepin both inhibited the response of SER-3⁺ cells to octopamine and the response to SER-4⁺ cells to serotonin (Fig. 2). The inhibitory effect of mianserin (but not methiothepin) was reversible within 5 min of removal (Fig. 2b, c). These results indicate that mianserin, an antidepressant that inhibits serotonin signalling in humans, extends *C. elegans* lifespan by blocking signalling through SER-4 serotonin receptors and SER-3 octopamine receptors (Fig. 2d, e).

Notably, 80–90% inhibition of the SER-3 response to octopamine required a tenfold higher concentration of mianserin than comparable inhibition of the SER-4 response to serotonin ($10 \mu\text{M}$ versus $1 \mu\text{M}$) (Fig. 2d, e). The difference was even greater for methiothepin. Thus, both drugs are more potent antagonists of SER-4 than SER-3. This suggests that exposure of animals to certain concentrations of these drugs might fully inhibit SER-4, but only partially inhibit SER-3, whereas higher concentrations might completely block both receptors. Interestingly, mianserin produced a large increase in lifespan at $50 \mu\text{M}$, but had little or no effect when used at $250 \mu\text{M}$ (Figs 1c and 3b). Together, these findings suggest that the lifespan-increasing effects of mianserin may result from a greater inhibition of SER-4 than SER-3, and that some activity of SER-3 may be necessary to achieve lifespan extension by mianserin.

Although a $50 \mu\text{M}$ concentration of mianserin increased lifespan in wild-type animals by $31 \pm 3\%$ when given only during adult life, it increased it by only $10 \pm 5\%$ when given during both larval and adult life (Fig. 3b and Table 1). No further lifespan increase was obtained by varying the drug's concentration, and a higher concentration caused developmental arrest (non-dauer) at the L2 larval stage (Fig. 3b and Supplementary Table 2). In addition, although *ser-3* and *ser-4* are required for lifespan increases in response to mianserin given to adults, lifespan increases in untreated mutants were only $6 \pm 3\%$ for *ser-3(ok2007)* and $19 \pm 2\%$ for *ser-4(ok512)* mutants, with the latter showing an 8% lower lifespan increase when it contained a wild-type *ser-4* transgene (Supplementary Table 3 and Supplementary Fig. 1). Thus, neither removing *ser-3* or *ser-4* nor blocking SER-3 and SER-4 receptors with mianserin during both

Table 1 | Effects of $50 \mu\text{M}$ mianserin on lifespan

Strain	Number of experiments	Mean lifespan (days) (+drug/–drug)	Percentage change	Number of animals (+drug/–drug)
N2/d1*	10	30.9/23.6	31 ± 3	1,180/1,915
N2/DR†	2	35.6/34.2	4 ± 2	229/248
N2/L1‡	5	26.6/24.1	10 ± 5	781/629
N2/d5§	4	22.3/22.1	1 ± 1	536/592
<i>clk-1(e2519)</i>	2	30.2/29.5	3 ± 4	241/239
<i>daf-16(mu86)</i>	3	19.7/17.2	14 ± 1	191/230
<i>daf-2(e1370)</i>	2	38.4/34.6	11 ± 4	302/236
<i>eat-2(ad1116)</i>	2	31.3/30.5	2 ± 5	192/292
<i>mod-1(ok103)</i>	2	32.0/25.7	25 ± 4	275/348
<i>mod-5(n3314)</i>	3	25.2/24.5	2 ± 9	217/236
<i>ser-1(ok345)</i>	2	30.2/22.2	38 ± 15	282/390
<i>ser-2(pk1357)</i>	2	28.3/23.8	19 ± 0	276/346
<i>ser-3(ok2007)</i>	3	26.2/26.8	-2 ± 2	385/537
<i>ser-3(ad1774)</i>	1	23.3/22.7	3	70/56
<i>ser-4(ok512)</i>	5	29.7/27.7	7 ± 2	677/746
<i>ser-7(tm1325)</i>	2	33.9/21.2	60 ± 11	381/358
<i>tph-1(mg280)</i>	3	25.1/25.0	0 ± 1	314/379

Cumulative statistics for mianserin treatment are shown. Mutant animals were given the drug starting at day 1 of adulthood. Values shown for mean lifespan and percentage change in lifespan are averages of the values obtained in individual experiments. For *P*-values and details see Supplementary Table 4a.

* Drug added to N2 animals on day 1 of adulthood.

† N2 animals subjected to dietary restriction.

‡ Drug added to N2 animals starting from the L1 stage.

§ Drug added to N2 animals on day 5 of adulthood.

larval and adult life recapitulates the lifespan-increasing effects of mianserin given only during adulthood. One possible explanation is that SER-3 and/or SER-4 loss of function in larvae induces compensatory mechanisms that alter the animal's physiology and, thereby, mianserin's effects on adults. Age-associated differences in loss of function phenotypes have previously been seen in mice²².

Does mianserin increase longevity via processes previously linked to ageing in *C. elegans*? We next tested mianserin on animals that have altered lifespans as a result of dietary restriction or because of mutations in genes encoding DAF-2 or DAF-16, two components of the IGF-1/insulin signalling pathway^{2,3}, CLK-1, a mitochondrial protein²³, or EAT-2, an ion channel subunit required for pumping food into the pharynx²⁴.

Our results indicate that lifespan extension by mianserin involves ageing mechanisms associated with dietary restriction. The combination of mianserin and dietary restriction increased lifespan only $4 \pm 2\%$ more than dietary restriction alone (Fig. 3c, d and Table 1). Furthermore, consistent with reported links between dietary restriction ageing mechanisms and increased lifespan in *eat-2(ad1116)* and

clk-1(e2519) mutants²⁴, mianserin increased the lifespan of these two mutants by only $2 \pm 5\%$ and $3 \pm 4\%$, respectively (Fig. 3c, d and Table 1). As in *tph-1(mg280)*, *mod-5(n3314)*, *ser-3(ok2007)* and *ser-4(ok512)* mutants, mianserin reduced serotonin-induced egg laying in both of these mutants, confirming drug uptake (Fig. 3a, left).

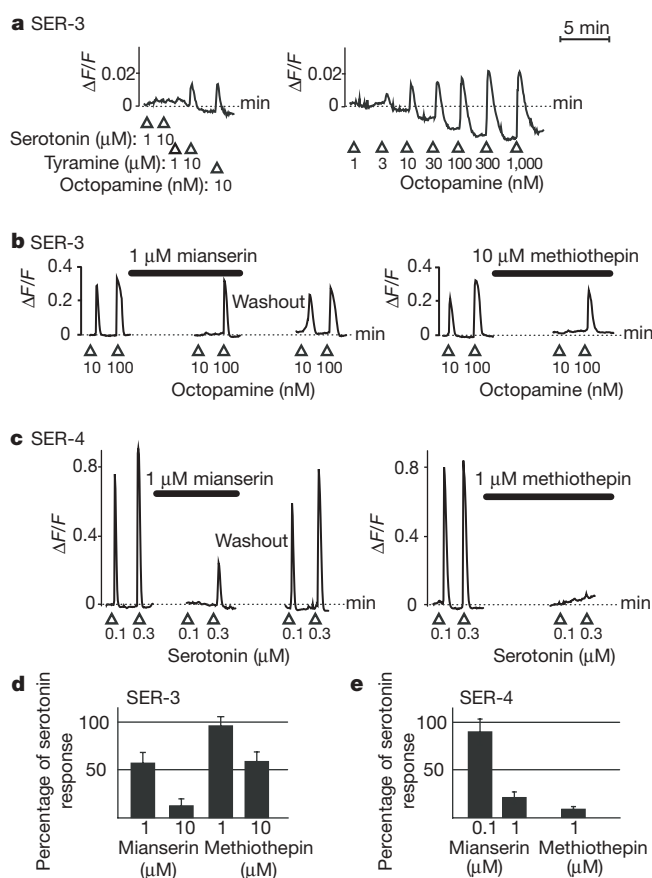


Figure 2 | Mianserin and methiothepin are antagonists of SER-3 octopamine and SER-4 serotonin receptors. HEK293 cells expressing SER-3 or SER-4 were monitored by calcium imaging for responses to potential ligands and drugs. In **a–c**, time is shown on the x axis and change in emitted fluorescence ($\Delta F/F$) on the y axis. Responses are shown for single cells in **b** and **c** and for groups of cells in **a**, **d** and **e**. **a**, SER-3⁺ cells showed calcium increases in response to 10 nM octopamine and 10 μM tyramine, but not to serotonin at 1 or 10 μM. EC₅₀ values for octopamine and tyramine were 24 nM and 26 μM, respectively. **b**, Mianserin and methiothepin (solid bars) inhibited the response of SER-3⁺ cells to 10 nM but not 100 nM octopamine, an effect that was reversible for mianserin. **c**, SER-4⁺ cells responded to 0.1 and 0.3 μM serotonin (EC₅₀ ~0.1 μM). Mianserin and methiothepin inhibited these responses, an effect that was reversible for mianserin. **d**, **e**, Bar graphs show the extent to which mianserin and methiothepin antagonized responses of SER-3⁺ cells to 10 nM octopamine (**d**) or SER-4⁺ cells to 300 nM serotonin (**e**). Responses are shown as percentages of responses seen in the absence of the drugs. Error bars indicate s.e.m.

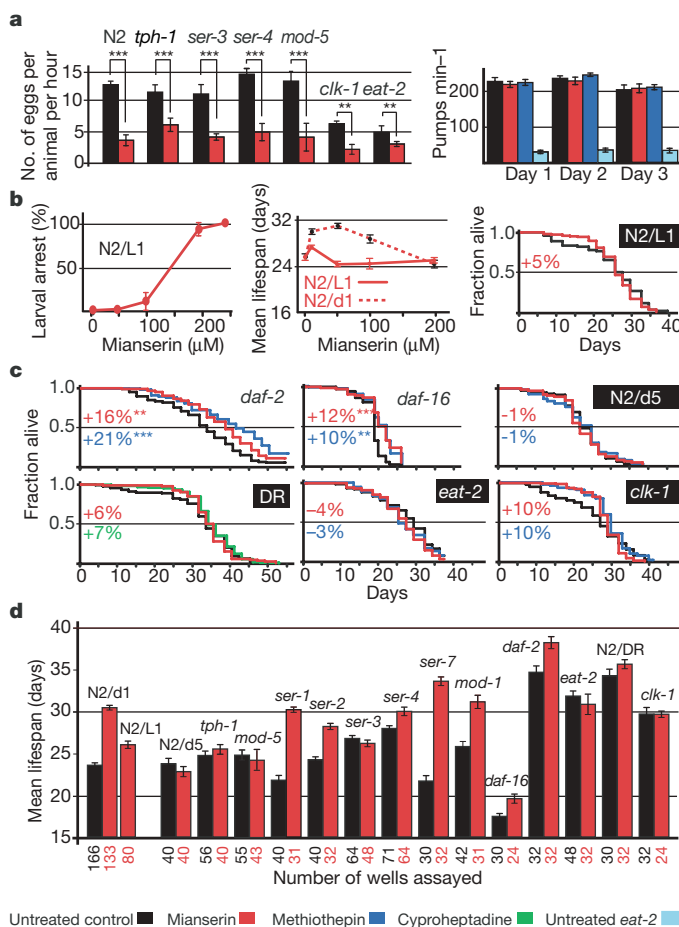


Figure 3 | Lifespan extension by mianserin and dietary restriction are related. Mianserin was tested for the ability to extend the lifespan of various ageing mutants or animals subjected to dietary restriction. **a**, Left: pre-exposure to mianserin reduced serotonin-induced egg laying even in strains in which it failed to increase lifespan. Animals were (red bars) or were not (black bars) pre-incubated with mianserin before exposure to serotonin (4–6 experiments per condition, double asterisk, $P \leq 0.001$, triple asterisk, $P \leq 0.0001$, one-way analysis of variance test). Right: pharyngeal pumps per minute on different days after receiving mianserin (red), methiothepin (blue), or control diluent (black), or in control *eat-2* mutants deficient in pumping (light blue). **b**, Mianserin had different effects when given to animals as both larvae and adults rather than as adults only. Left: exposure to increasing concentrations of mianserin starting from the L1 stage led to progressive larval arrest. Middle: animals that did reach adulthood when exposed to mianserin as larvae (solid red line, N2/L1) showed little increase in lifespan in response to mianserin exposure as adults compared to animals exposed to the drug only as adults (dotted red line, N2/d1). Mean lifespan in days is shown as a function of increasing mianserin concentration for animals that reached adulthood. Right: survival curve for animals exposed to 50 μM mianserin starting from the larval L1 stage. All controls were assayed in parallel. Animals in which mianserin failed to increase lifespan are highlighted in black in **b** and **c**. **c**, Survival curves representing a typical experiment show the fraction of animals alive at different adult ages (in days) when animals were exposed to mianserin (red), methiothepin (blue), or control diluent (black). Percentage change and P values in **b** and **c** are indicated as in Fig. 1. (Sample sizes, ≥ 50). DR, dietary restriction. **d**, Mean lifespan in days for mianserin-treated and untreated control animals (cumulative). Mean lifespan and s.e.m. were calculated by averaging the lifespans of identically treated wells (10–15 animals per well). The number of wells assayed for each condition is indicated underneath each bar. Error bars indicate s.e.m. See Supplementary Tables 2–4 for details.

Although serotonin is involved in pharyngeal pumping^{11,25}, two findings indicate that mianserin does not reduce food intake. First, although *eat-2(ad1116)* mutants showed an 80–90% decrease in pharyngeal pumping, pumping rates were the same in mianserin-treated and untreated animals (Fig. 3a). Second, dietary restriction can increase lifespan when initiated as late as day 10 of adulthood²⁶, whereas exposure to mianserin beginning on day 5 had no effect on lifespan (Fig. 3c, d and Table 1). These results indicate that, although mianserin extends lifespan through mechanisms associated with dietary restriction, it does not do so by decreasing food intake.

Mianserin increased the lifespan of *daf-2(e1370)* and *daf-16(mu86)* mutants by $11 \pm 4\%$ and $14 \pm 1\%$, respectively, suggesting that the two genes might also influence mianserin-induced lifespan extension (Fig. 3c, d). However, these results do not compare to the marked results obtained with dietary-restricted animals or *eat-2(ad1116)* and *clk-1(e2519)* mutants. These findings contrast with a requirement for *daf-16* for increased reproductive period (reproductive longevity) in *tph-1(mg280)* mutants. In our studies, untreated *tph-1(mg280)* mutants showed only a $4 \pm 5\%$ increase in lifespan compared to wild-type animals, further distinguishing the effects of serotonin signalling on reproductive longevity versus lifespan¹³.

We found that *C. elegans* lifespan is increased by giving adult animals mianserin, a drug used as an antidepressant in humans. This effect requires the presence of serotonin as well as two neurotransmitter receptors: the SER-4 serotonin receptor and the SER-3 octopamine receptor. Similar to its antagonistic action on human serotonin receptors, mianserin inhibits both SER-4 and SER-3. Serotonin and octopamine are thought to serve as physiological antagonists that signal the presence of food (serotonin) versus starvation (octopamine) in *C. elegans*^{11,16,25}. It may be that these two neurotransmitters exist in a dynamic equilibrium that is tipped in the direction of a starvation response by mianserin, possibly because of the greater inhibitory effect of mianserin on SER-4 than SER-3. In this way, mianserin might potentially create a 'perceived' state of starvation that, despite adequate food intake, would activate mechanisms of lifespan extension downstream of dietary restriction. Interestingly, one side effect of mianserin in humans is increased appetite, suggesting a possible evolutionary link between appetite and lifespan in *C. elegans* and humans²⁷.

METHODS SUMMARY

Lifespan assay. Lifespan was assessed in liquid medium^{5,28} at 20 °C in 384- or 96-well plates. Age-synchronized *C. elegans* were distributed (seeded) in wells as L1 larvae (10–15 (96-well plates) or 6–12 (384-well plates) animals per well) together with *Escherichia coli* OP50. To prevent self-fertilization, 5-fluoro-2'-deoxyuridine (Sigma) was added 42–45 h after seeding (0.12 mM final). Unless otherwise specified, drugs were added 68 h after seeding, which corresponded to day 1 of adult life and of the lifespan assay. The fraction of animals alive was scored on the basis of body movement.

Statistical analysis. Comparisons and *P*-value calculations were made between treated and untreated animals of the same strain using the log-rank test (Mantel-Haenszel).

Calcium imaging. HEK293 cells were transfected with expression plasmids²¹ encoding SER-4 or Gα15 and SER-3. Potential ligands were applied for 4 s, with 2 min separating different applications. Fluorescence emission was determined every 4 s. To test the effects of mianserin and methiothepin, cells were first exposed to a ligand and responses were recorded. Drugs were then applied for 5 min, after which responses to the ligands were tested again. Inhibition of responses by mianserin or methiothepin was calculated by the following equation: percentage inhibition = $100 \times ((\text{response in the presence of inhibitor})/(\text{uninhibited response}))$.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 31 July; accepted 11 October 2007.

1. Kirkwood, T. B. Understanding the odd science of aging. *Cell* **120**, 437–447 (2005).

2. Kenyon, C. The plasticity of aging: insights from long-lived mutants. *Cell* **120**, 449–460 (2005).
3. Finch, C. E. & Ruvkun, G. The genetics of aging. *Annu. Rev. Genomics Hum. Genet.* **2**, 435–462 (2001).
4. Evason, K., Huang, C., Yamben, I., Covey, D. F. & Kornfeld, K. Anticonvulsant medications extend worm life-span. *Science* **307**, 258–262 (2005).
5. Melov, S. *et al.* Extension of life-span with superoxide dismutase/catalase mimetics. *Science* **289**, 1567–1569 (2000).
6. Kang, H. L., Benzer, S. & Min, K. T. Life extension in *Drosophila* by feeding a drug. *Proc. Natl Acad. Sci. USA* **99**, 838–843 (2002).
7. Wood, J. G. *et al.* Sirtuin activators mimic caloric restriction and delay ageing in metazoans. *Nature* **430**, 686–689 (2004).
8. Terzibasi, E., Valenzano, D. R. & Cellerino, A. The short-lived fish *Nothobranchius furzeri* as a new model system for aging studies. *Exp. Gerontol.* **42**, 81–89 (2007).
9. Gillman, P. K. A systematic review of the serotonergic effects of mirtazapine in humans: implications for its dual action status. *Hum. Psychopharmacol.* **21**, 117–125 (2006).
10. Dempsey, C. M., Mackenzie, S. M., Gargus, A., Blanco, G. & Sze, J. Y. Serotonin (5HT), fluoxetine, imipramine and dopamine target distinct 5HT receptor signaling to modulate *Caenorhabditis elegans* egg-laying behavior. *Genetics* **169**, 1425–1436 (2005).
11. Horvitz, H. R., Chalfie, M., Trent, C., Sulston, J. E. & Evans, P. D. Serotonin and octopamine in the nematode *Caenorhabditis elegans*. *Science* **216**, 1012–1014 (1982).
12. Rodin, G. *et al.* Treatment of depression in cancer patients. *Curr. Oncol.* **14**, 180–188 (2004).
13. Sze, J. Y., Victor, M., Loer, C., Shi, Y. & Ruvkun, G. Food and metabolic signalling defects in a *Caenorhabditis elegans* serotonin-synthesis mutant. *Nature* **403**, 560–564 (2000).
14. Ranganathan, R., Sawin, E. R., Trent, C. & Horvitz, H. R. Mutations in the *Caenorhabditis elegans* serotonin reuptake transporter MOD-5 reveal serotonin-dependent and -independent activities of fluoxetine. *J. Neurosci.* **21**, 5871–5884 (2001).
15. Komuniecki, R. W., Hobson, R. J., Rex, E. B., Hapiak, V. M. & Komuniecki, P. R. Biogenic amine receptors in parasitic nematodes: what can be learned from *Caenorhabditis elegans*? *Mol. Biochem. Parasitol.* **137**, 1–11 (2004).
16. Suo, S., Kimura, Y. & Van Tol, H. H. Starvation induces cAMP response element-binding protein-dependent gene expression through octopamine-Gq signaling in *Caenorhabditis elegans*. *J. Neurosci.* **26**, 10082–10090 (2006).
17. Murakami, H. & Murakami, S. Serotonin receptors antagonistically modulate *Caenorhabditis elegans* longevity. *Aging Cell* **6**, 483–488 (2007).
18. Roeder, T. Octopamine in invertebrates. *Prog. Neurobiol.* **59**, 533–561 (1999).
19. Malnic, B., Hirono, J., Sato, T. & Buck, L. B. Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999).
20. Liberles, S. D. & Buck, L. B. A second class of chemosensory receptors in the olfactory epithelium. *Nature* **442**, 645–650 (2006).
21. Offermanns, S. & Simon, M. I. Gα15 and Gα16 couple a wide variety of receptors to phospholipase C. *J. Biol. Chem.* **270**, 15175–15180 (1995).
22. Luquet, S., Perez, F. A., Hnasko, T. S. & Palmiter, R. D. NPY/AgRP neurons are essential for feeding in adult mice but can be ablated in neonates. *Science* **310**, 683–685 (2005).
23. Lakowski, B. & Hekimi, S. Determination of life-span in *Caenorhabditis elegans* by four clock genes. *Science* **272**, 1010–1013 (1996).
24. Lakowski, B. & Hekimi, S. The genetics of caloric restriction in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **95**, 13091–13096 (1998).
25. Niacaris, T. & Avery, L. Serotonin regulates repolarization of the *C. elegans* pharyngeal muscle. *J. Exp. Biol.* **206**, 223–231 (2003).
26. Kaeberlein, T. L. *et al.* Lifespan extension in *Caenorhabditis elegans* by complete removal of food. *Aging Cell* **5**, 487–494 (2006).
27. Harris, B. & Harper, M. Unusual appetites in patients on mianserin. *Lancet* **1**, 590 (1980).
28. Johnson, T. E. *et al.* Relationship between increased longevity and stress resistance as assessed through gerontogene mutations in *Caenorhabditis elegans*. *Exp. Gerontol.* **36**, 1609–1617 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to J. Priess for critical reading of an earlier version of the manuscript, to members of the Priess and Buck laboratories for advice and discussions, and to J. Vazquez for technical assistance. We thank S. Suo for providing the VN11 strain (*ser-3(ad1774);tZls3[cre::gfp, lin-15(+)]*) and J. Ying Sze for the *ser-4(ok512);yZEx205[ser-4(+); pRF4(rol-6(su1006))]* strain. All other strains were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH, and the international *C. elegans* Gene Knockout Consortium. This project was supported by the Howard Hughes Medical Institute and the Ellison Medical Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.B.B. (lbuck@fhcc.org).

METHODS

Strains and genetics. All strains were maintained at 20 °C as described in ref. 29. *Caenorhabditis elegans* strains used were: Bristol strain (N2), CB4876 *clk-1(e2519)* III, GR1321 *tph-1(mg280)* II, CF1038 *daf-16(mu86)* I, CB1370 *daf-2(e1370)* III, DA1116 *eat-2(ad1116)* II, AQ866 *ser-4(ok512)* III, *ser-4(ok512);yzEx205[ser-4(+); pRF4(rol-6(su1006))]*, OHO313 *ser-2(pk1357)* X, DA1814 *ser-1(ok345)* X, MT9668 *mod-1(ok103)* V, MT9772 *mod-5(n3314)* I, CB1370 *daf-2(e1370)* III, DA2100 *ser-7(tm1325)* X, RB1631 *ser-3(ok2007)* I, and VN11 *ser-3(ad1774);tZIs3[cre::gfp, lin-15(+)]*.

Lifespan assay. Lifespans were assessed in liquid medium^{5,28} (S-complete medium with 50 µg ml⁻¹ carbenicillin and 0.1 µg ml⁻¹ fungizone) in 384- or 96-well plates containing, respectively, 60 µl or 150 µl total volume, 6–12 or 10–15 nematodes, and 9 mg ml⁻¹ or 6 mg ml⁻¹ freshly prepared *E. coli* OP50 per well. Age-synchronized nematodes were seeded as L1 larvae and plates were sealed with tape (Nunc) to prevent evaporation. To prevent self-fertilization, 5-fluoro-2'-deoxyuridine (0.12 mM final) (Sigma) was added 42–45 h after seeding. Drugs were added 68 h after seeding (day 1 of adult life) unless otherwise specified. Day 1 of the lifespan assay started 68 h after seeding the animals into plates.

The fraction of animals alive per well was scored microscopically on the basis of movement. Before counting, each plate was put onto a plate rotator for 1–2 min. Strong microscope light (visual or ultraviolet) effectively stimulated movement even in old animals. Using this assay, *daf-16*, *eat-2* and *clk-1* ageing mutants showed alterations in lifespan similar to those reported using standard conditions (agar plates)^{2,3,24}.

Dietary restriction. Animals were grown in liquid medium in the presence of food until day 3 of adulthood, when food was reduced 25-fold by serial dilution²⁶.

Statistical analysis. STATA8 software was used for analysis. Comparisons and *P* value calculations were made between treated and untreated animals of the same strain using the log-rank test (Mantel–Haenszel). We observed the death of 98.6% of the animals (excluding screen). Animals that were still alive at the end of an experiment (1.4%) were analysed as alive up to this point with unknown time of death (censoring). Wells containing more than 19 animals were excluded from the analysis.

Pharyngeal pumping. Animals were grown in liquid medium at 20 °C (with or without drug). After the animals were transferred to bacteria-coated agar plates (with or without drug) and then left for 30 min, the grinder movements within a 10-s interval were counted. Animals were assayed on three consecutive days after drug treatment. Numbers of animals tested on the three days were: controls, *n* = 67/67/58; mianserin, *n* = 57/68/57; methiothepin, *n* = 26/27/24.

Egg-laying assay. The egg-laying assay was performed as described previously¹⁰ but using 50 µM mianserin instead of 20 µM. Numbers of animals tested (with/without mianserin): N2 (64/54); *clk-1(e2519)* (26/26); *eat-2(ad1116)* (38/29); *mod-5(n3314)* (40/40); *ser-3(ok2007)* (39/29); *ser-4(ok512)* (48/38); *tph-1(mg280)* (47/40).

Expression vectors. The *ser-3* or *ser-4* coding region was amplified by PCR from cDNA prepared from *C. elegans* RNA, and then cloned into the pcDNA3.1(-) expression vector (Invitrogen) to give the SER-3 (pMP513#6) or SER-4 (pMP509#6) expression vector. In pMP509#6, sequence encoding the first 20 amino acids of bovine rhodopsin was added to the 5' end of the *ser-4* coding region to potentially enhance cell surface expression²⁰.

Calcium imaging. A total of 4×10^5 HEK293 cells were seeded into individual wells of 6-well plates containing coverslips coated with poly D-lysine, and then transfected with pMP509#6 (SER-4, 200 ng per well) and a Gα15 expression plasmid²¹ (150 ng per well), or with pMP513#6 (SER-3, 200 ng per well), using lipofectamine according to the manufacturer's instructions (Invitrogen, catalogue number 1514-015). After 24 h, cells were loaded with the calcium indicator calcium 3 (Molecular Devices) in HBSS/20 mM HEPES (Gibco) for 1 h before imaging. Calcium imaging was done on coverslips in a perfusion chamber mounted on an inverted microscope (Olympus Ix70) using a 10×/0.3 NA objective (Olympus UplanFI) to maximize the number of imaged neurons. During imaging, cells were continuously perfused with HBSS and intermittently exposed to HBSS containing ligands and/or drugs. Ligands were applied for 4 s, with 2 min separating different applications. Fluorescence emission was determined every 4 s using a CCD camera (Hamatsu C4742-95-10NR) and a standard filter set (high Q filter set (R.P.I.): 470/40 excitation filter; 495 nm low-pass filter dichroic mirror; 525/50 nm emissions filter). To test the drugs mianserin and methiothepin, cells were first exposed to a ligand and responses were recorded. Drugs were added to the perfusion buffer and continuously applied for 5 min, after which responses to the ligands were tested again. The perfusion buffer was changed back to pure HBSS for a 5 min washout, after which ligands were applied again. Image analysis was done using Metafluor software (Molecular Devices). Fluorescent signals were normalized using the following equation: $\Delta F/F = (F_t - F_0)/F_0$. Inhibition of responses by mianserin or methiothepin was calculated by the following equation: percentage inhibition = $100 \times ((\text{response in the presence of inhibitor})/(\text{uninhibited response}))$.

Drug preparation. Mianserin, methiothepin and cyproheptadine (Sigma) were freshly prepared each time and dissolved in water at 50× final concentration before use.

29. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).

Social evaluation by preverbal infants

J. Kiley Hamlin¹, Karen Wynn¹ & Paul Bloom¹

The capacity to evaluate other people is essential for navigating the social world. Humans must be able to assess the actions and intentions of the people around them, and make accurate decisions about who is friend and who is foe, who is an appropriate social partner and who is not. Indeed, all social animals benefit from the capacity to identify individual conspecifics that may help them, and to distinguish these individuals from others that may harm them. Human adults evaluate people rapidly and automatically on the basis of both behaviour and physical features^{1–6}, but the ontogenetic origins and development of this capacity are not well understood. Here we show that 6- and 10-month-old infants take into account an individual's actions towards others in evaluating that individual as appealing or aversive: infants prefer an individual who helps another to one who hinders another, prefer a helping individual to a neutral individual, and prefer a neutral individual to a hindering individual. These findings constitute evidence that preverbal infants assess individuals on the basis of their behaviour towards others. This capacity may serve as the foundation for moral thought and action, and its early developmental emergence supports the view that social evaluation is a biological adaptation.

Our experiments used two methodologies to assess 6- and 10-month-old infants' intuitions about social interactions: a choice paradigm^{7–11} in which infants indicate preferences through their reaching behaviour, and a violation of expectation paradigm^{12–16} that assesses infants' expectations via their looking times, capitalizing on the phenomenon that infants tend to look longer at unexpected or surprising events.

In experiment 1, infants saw a character (the 'climber', made of wood and with large eyes glued onto it) initially at rest at the bottom of a hill. During a habituation phase, infants saw events in which the climber repeatedly attempted to climb the hill, and on the third attempt was either aided up by a helper who pushed it from behind, or was pushed down by a hinderer (Fig. 1a; stimulus clips and Supplementary Methods are available at <http://www.yale.edu/infantlab/socialevaluation> and in the Supplementary Information). Infants saw alternating helping and hindering trials with looking time measured on each trial, until their looking time reached a pre-set criterion indicating they had sufficiently processed these events. Previous studies from our laboratory found that infants interpret similar, computer-animated events as instances of helping and hindering, and expect the climber subsequently to approach the helper and avoid the hinderer^{15,16}. We asked here how infants themselves, as uninvolved and unaffected bystanders, evaluate helpers and hinderers. Will witnessing one individual's actions towards an unknown third party affect infants' attitude towards that individual?

In the test phase, our choice measure examined infants' attitudes towards the helper and hinderer. Infants were encouraged to choose between the two (that is, reach for one). Infants robustly chose the helper (14 of 16 10-month-olds, binomial probability test, one-tailed $P = 0.002$; 12 of 12 6-month-olds, $P = 0.0002$), indicating that they

held distinct impressions of the two characters on the basis of their actions towards the climber (see Fig 2).

Our looking time measure replicated our previous studies assessing 9- and 12-month-olds' expectations about the climber's attitudes to the helper and hinderer^{15,16}, and extended this question to younger infants. Infants saw a new display containing climber, helper and hinderer (Fig. 1b). The climber alternately approached the helper (unsurprising) and the hinderer (a surprising action). Replicating our previous results, 10-month-olds looked longer at the latter event (mean_{hinderer} = 4.96 s, mean_{helper} = 3.82 s; paired t -test, $t(15) = 2.603$, two-tailed $P = 0.02$), indicating surprise when the climber approached one who had previously hindered it. Six-month-olds, however, looked equally to both events (mean_{hinderer} = 5.7 s, mean_{helper} = 6.7 s; $t(11) = 0.80$, $P = 0.44$), suggesting that they did not attribute to the climber distinct attitudes towards the two characters, despite themselves preferring helper to hinderer in our choice measure. This suggests that the capacity for social evaluation may develop before the ability to infer others' evaluations.

Our claim—that young infants evaluate others based on their social behaviour—entails that infants were responding to social, not superficial perceptual, aspects of our events. If infants of these ages prefer, for example, upward to downward motion, or pushing up to pushing down actions, then our subjects may have chosen the helper for these non-social reasons. To assess this, we conducted a second experiment in which infants saw events like those of experiment 1 except that the pushed object did not appear animate and goal-driven like our climber, but inanimate (Fig. 1c)—an entity to which social notions of helping and hindering do not apply.

In experiment 2, new groups of infants saw two characters (the helper and hinderer from experiment 1), on alternating trials, smoothly push up or down the hill, respectively, an inanimate object (the climber of experiment 1 but with eyes removed, and undergoing no self-propelled motion). Although the two characters' physical trajectories and respective effects on the pushed object were similar to those of the helper and hinderer in experiment 1, these events are not social interactions and cannot be viewed as instances of 'helping' and 'hindering'. Infants were then given a choice measure only.

If perceptual preferences, not social evaluations, drove infants' choices in experiment 1, similar preferences should be obtained here: infants should robustly prefer the pusher-up to the pusher-down character. However, neither age group did so. Six of twelve 10-month-olds chose the pusher-up, binomial probability test, one-tailed $P = 0.613$; four of twelve 6-month-olds did so, $P = 0.927$. These responses differed significantly from those in experiment 1 (10-month-olds, $P = 0.04$, Fisher's exact test; 6-month-olds, $P = 0.001$), in which infants overwhelmingly chose the pusher-up (helper) when this action was embedded in a social context, indicating that infants' preferences in experiment 1 were based on social, not perceptual, differences between helping and hindering events.

Infants' choice patterns indicate three possibilities: infants may positively evaluate an individual seen helping another (thus find the helper appealing); they may negatively evaluate an individual

¹Yale University, Department of Psychology, New Haven, Connecticut 06520-8205, USA.

seen hindering another (thus find the hinderer aversive); or both positive and negative evaluation processes may be operative. We accordingly conducted a third experiment in which new groups of 6- and 10-month-old infants chose between a neutral character and a valenced character—either a helper (for infants in the helping/neutral condition) or a hinderer (the hindering/neutral condition).

In experiment 3 habituation trials, each infant saw either a helper or a hinderer acting on a climber as in experiment 1, and a neutral character which moved uphill or downhill in the same manner as the valenced character but did not interact with the climber (Fig. 1d). Infants in both conditions were then given a choice measure to assess their own preference for the neutral versus valenced character, and a looking time measure to assess their expectations of the climber's preference for the neutral versus valenced character.

In the choice measure, infants of both ages responded differently to the neutral character when it was paired with the helper than when it was paired with the hinderer (Fisher's exact test, two-tailed $P = 0.01$ for each age group). Infants in the helping/neutral condition systematically chose the helper (seven of eight 10-month-olds, binomial probability test, one-tailed $P = 0.035$; and seven of eight 6-month-olds, $P = 0.035$), whereas infants in the hindering/neutral group chose the neutral character (seven of eight 10-month-olds, $P = 0.035$; and seven of eight 6-month-olds, $P = 0.035$). That is, infants were both drawn towards helpers and independently inclined to avoid hinderers, revealing both positive and negative evaluations. Infants' choices were not based on general perceptual preferences: within each condition, the neutral and valenced characters enacted identical motion patterns; a preference for solitary action over interaction (or vice versa) would have generated across-the-board choice (or avoidance) of the neutral character in both conditions, not choice in one and avoidance in the other, as was observed. Infants'

preference for the helper and aversion to the hinderer, then, are best explained as specifically social evaluations: a liking for those who act cooperatively to facilitate the goals of others, and a dislike of those who impede another's goals.

In our looking time measure, infants of both ages failed to discriminate the test events (climber approaching the neutral versus the valenced character (helper or hinderer), two-tailed t -tests, all P -values > 0.3). Together with the looking time results from experiment 1 this suggests that 10-month-olds, although having expectations of how an individual will respond to two actors performing opposing actions, do not anticipate how an individual will respond to actors performing less-distinct actions.

Previous research has established that infants in the first half-year of life exhibit preferences for social individuals based on static perceptual features (for example, facial attractiveness, race)^{17,18}, and toddlers by 18 months of age spontaneously engage in cooperative helping behaviour¹⁹. The findings reported here constitute the first evidence that young infants' social preferences are influenced by others' behaviour towards unrelated third parties. The presence of social evaluation so early in infancy suggests that assessing individuals by the nature of their interactions with others is central to processing the social world, both evolutionarily and developmentally. The capacity for such evaluations can be seen as a biological adaptation: cooperative behaviour such as group hunting, food sharing and warfare can be beneficial to individual members of a group, but can only successfully evolve if individuals can distinguish free riders from cooperators or 'reciprocators', those willing to do their fair share^{20,21}. Our findings suggest that preverbal infants may be sensitive to this distinction.

The capacity to evaluate individuals by their social actions may also serve as a foundation for a developing system of moral cognition. Plainly, many aspects of a full-fledged moral system are beyond the

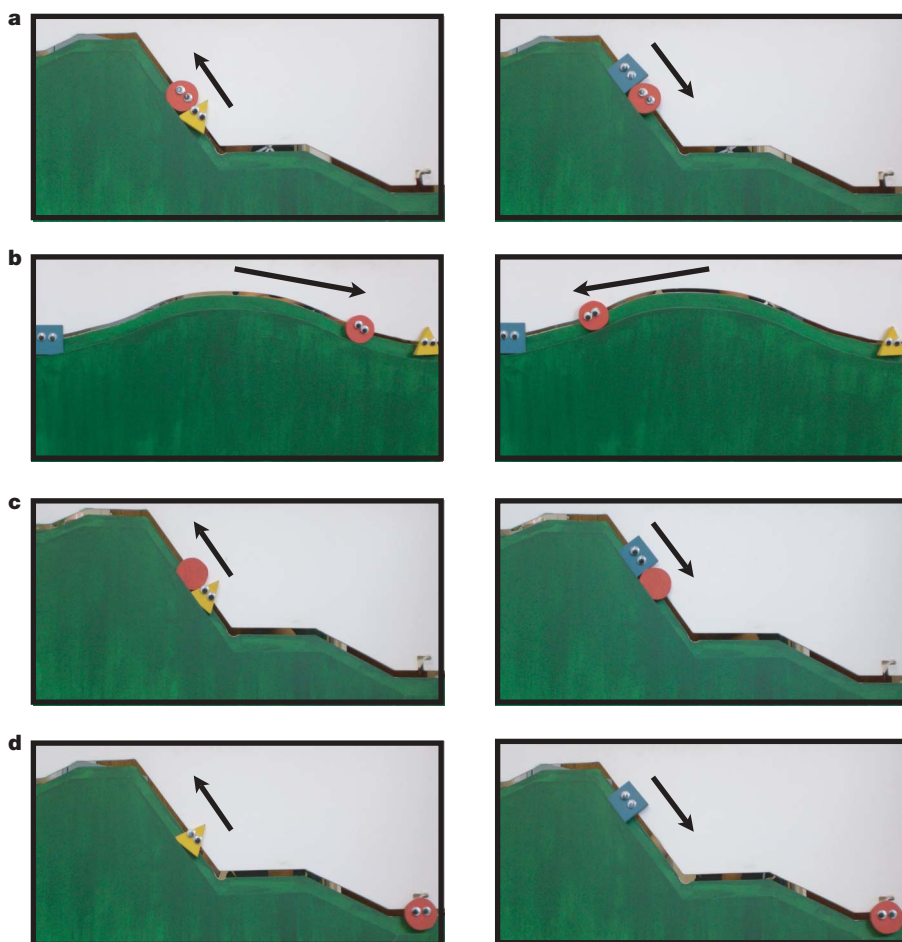


Figure 1 | Social interaction events shown to infants. **a**, Helping and hindering habituation events of experiments 1 and 3. On each trial, the climber (red circle) attempts to climb the hill twice, each time falling back to the bottom of the hill. On the third attempt, the climber is either bumped up the hill by the helper (left panel) or bumped down the hill by the hinderer (right panel). Infants in experiment 1 saw these two events in alternating sequence; infants in experiment 3 saw either a helping or hindering event in alternation with the corresponding neutral event depicted in **d**. **b**, Looking time test events of experiments 1 and 3. The climber moves from the top of the hill to sit with the character on the right (left panel) or the left (right panel). **c**, Pushing-up and pushing-down habituation events of experiment 2. An inanimate object (red circle) rests (left panel) at the bottom of the hill and is pushed up, or rests (right panel) at the top of the hill and is pushed down. Infants saw these two events in alternation. **d**, Neutral habituation events from helper/neutral (left panel) and hinderer/neutral (right panel) conditions of experiment 3. The neutral character, without interacting with the climber, traces a path identical to that of the helper (left panel) or hinderer (right panel). Each infant saw either the helping or hindering event depicted in **a**, in alternation with the corresponding neutral event.

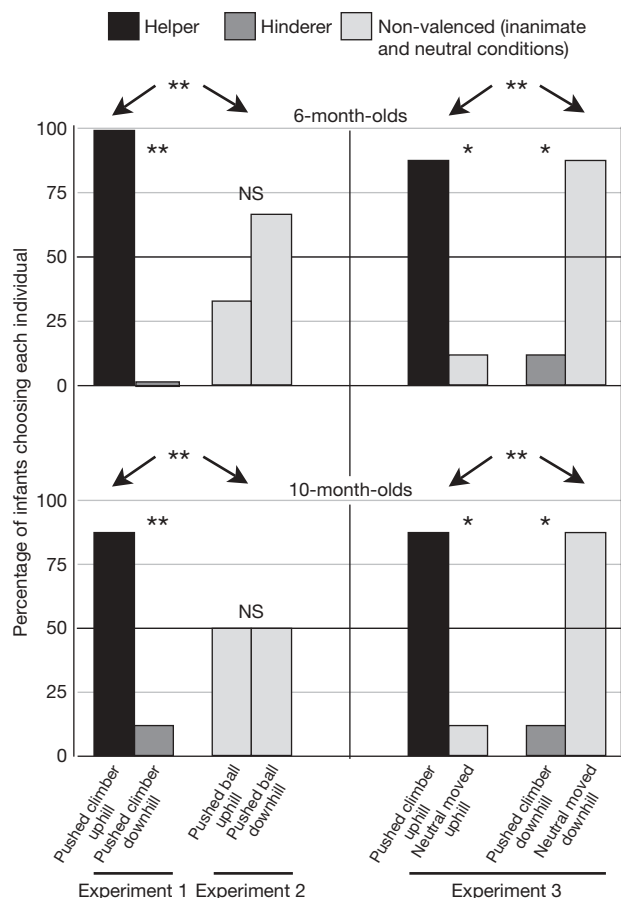


Figure 2 | Choice results. Percentage of infants choosing each character across experiments 1, 2 and 3. NS, not significant. Asterisk, one-tailed $P < 0.05$; double asterisk, $P < 0.05$.

grasp of the preverbal infant. Yet the ability to judge differentially those who perform positive and negative social acts may form an essential basis for any system that will eventually contain more abstract concepts of right and wrong. The social evaluations we have observed in our young subjects have (at least) one crucial component of genuine moral judgements: they do not stem from infants' own experiences with the actors involved. Our subjects had no previous history with our characters, nor did they themselves experience any consequences of these characters' actions. Their evaluations were made on the basis of witnessed interactions between unknown individuals: the infant, as an unaffected, unrelated (and therefore unbiased) third party, is nonetheless rendering a judgement about the value of a social act.

Our findings indicate that humans engage in social evaluation far earlier in development than previously thought, and support the view that the capacity to evaluate individuals on the basis of their social interactions is universal and unlearned^{22–24}. Determining the complexity of this understanding—for instance, do infants prefer to interact with agents who punish hinderers over those who reward them—will require further research.

METHODS SUMMARY

Subjects were healthy full-term infants recruited from the greater New Haven area and tested in K.W.'s Infant Cognition Laboratory at Yale University. Ten-month-olds ranged from 9 months 12 days to 10 months 16 days; 6-month-olds from 5 months 3 days to 6 months 17 days.

Habituation events occurred in a display (122 cm wide, 66 cm high) containing a green incline with 43-cm elevation from base to top. Characters were blocks (9 cm × 9 cm × 1 cm) with large (2.5-cm diameter) 'googly eyes' (with the exception of the eyeless object in experiment 2). The looking time test display (122 cm × 66 cm) contained a hill with 14-cm elevation from base to top.

Infants sat in parents' laps; parents were instructed not to interfere with infants. Parents of all infants in experiment 2 and 6-month-olds in experiment 3 were additionally instructed to close their eyes during choice measure. Infants received habituation trials until (1) looking time on three consecutive trials (after the first three) decreased to half that on the first three trials or (2) 14 trials were presented. End-of-trial for habituation and looking time test trials occurred when (1) the infant looked away continuously for 2 s or (2) 60 s had elapsed. A coder blind to the identities of the characters monitored infants' looking times and administered the choice measure. A second coder independently coded a random 25% of subjects of each age group in each experiment; coders achieved 98% positive agreement on both measures.

The following were counterbalanced across subjects in each experiment and age group: identities of helper/hinderer (experiment 1), pusher-up/pusher-down (experiment 2) and valenced/neutral characters (experiment 3); order of habituation events; order of choice and looking time measures (experiments 1 and 3); positions (left/right) of characters in choice and in looking time trials; order of climber's approach in looking time test trials to helper/hinderer (experiment 1) and valenced/neutral character (experiment 3).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 August; accepted 24 September 2007.

- Langlois, J. H. *Physical Appearance, Stigma, and Social Behavior: The Ontario Symposium* Vol. 3 (eds Herman, C. P., Zanna, M. P. & Higgins, E. T.) (Erlbaum, Hillsdale, New Jersey, 1986).
- Shepherd, J. *Perceiving and Remembering Faces* (eds Davies, G., Ellis, H. & Shepherd, J.) (Academic Press, San Diego, California, 1981).
- Winter, L. & Uleman, J. When are social judgments made? Evidence for the spontaneousness of trait inferences. *J. Pers. Soc. Psychol.* **47**, 237–252 (1984).
- Ambady, N. & Rosenthal, R. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychol. Bull.* **111**, 256–274 (1992).
- Cunningham, W. A. et al. Separable neural components in the processing of black and white faces. *Psychol. Sci.* **15**, 806–813 (2004).
- Todorov, A., Mandisodza, A., Goren, A. & Hall, C. Inferences of competence from faces predict election outcomes. *Science* **308**, 1623–1626 (2005).
- Piaget, J. *The Construction of Reality in the Child* (Routledge, New York, 1954).
- Repacholi, B. M. & Gopnik, A. Early reasoning about desires: Evidence from 14–18-month-olds. *Dev. Psychol.* **33**, 12–21 (1997).
- Feigenson, L., Carey, S. & Spelke, E. S. Infants' discrimination of number vs continuous extent. *Cognit. Psychol.* **44**, 33–66 (2002).
- Hamlin, J. K., Hallinan, E. V. & Woodward, A. L. Do as I do: 7-month-old infants selectively reproduce others' goals. *Dev. Sci.* (in the press).
- Cherries, E., Mitroff, S. R., Wynn, K. & Scholl, B. J. Cohesion as a principle of object persistence in infancy. *Dev. Sci.* (in the press).
- Woodward, A. L. Infants selectively encode the goal object of an actor's reach. *Cognition* **69**, 1–34 (1998).
- Premack, D. & Premack, A. J. Infants attribute value to the goal-directed actions of self-propelled objects. *J. Cogn. Neurosci.* **9**, 848–856 (1997).
- Gergely, G., Nadasdy, Z., Csibra, G. & Biro, S. Taking the intentional stance at 12 months of age. *Cognition* **56**, 165–193 (1995).
- Kuhlmeier, V., Wynn, K. & Bloom, P. Attribution of dispositional states by 12-month-olds. *Psychol. Sci.* **14**, 402–408 (2003).
- Kuhlmeier, V., Wynn, K. & Bloom, P. Reasoning about present dispositions based on past interactions (International Conference on Infant Studies, Chicago, Illinois, 5 May 2004).
- Kelley, D. et al. 3-month-olds, but not newborns, prefer own-race faces. *Dev. Sci.* **8**, F31–F36 (2005).
- Slater, A. et al. Newborn infants prefer attractive faces. *Infant Behav. Dev.* **21**, 345–354 (1998).
- Warneken, F. & Tomasello, M. Altruistic helping in human infants and young chimpanzees. *Science* **311**, 1301–1303 (2006).
- Trivers, R. L. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57 (1971).
- Axelrod, R. *The Evolution of Cooperation* (Basic Books, New York, 1984).
- Haidt, J. & Joseph, C. Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus* **133**, 55–66 (2004).
- Pinker, S. *The Blank Slate* (Norton, New York, 2002).
- Hauser, M. *Moral Minds* (Ecco, New York, 2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Norman, J. Stitelman, E. Madva, K. McCrink, G. Newman and E. Cherries for their assistance and input. This work was supported in part by a National Science Foundation grant to K.W.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.K.H. (kiley.hamlin@yale.edu) and to K.W. (karen.wynn@yale.edu).

METHODS

Experiment 1: helper versus hinderer. Subjects were 16 10-month-olds (8 girls; mean age 9 months, 26 days; range 9 months, 14 days to 10 months, 16 days) and 12 6-month-olds (5 girls; mean age 6 months, 5 days; range 5 months, 18 days to 6 months, 17 days). Four additional infants (two 10-month-olds) were excluded owing to parental interference (two) or procedure error (two). Subjects in all experiments were healthy full-term infants.

Infants sat in parents' laps before a table with a curtain at the far end (165 cm from the infant), which could be lowered to occlude a display stage (122 cm wide, 66 cm high) with a white background and a green 'hill' or incline protruding 10 cm, which rose from lower right to upper left corner of the display (43 cm from lowest to highest point). It had a small 'plateau' one-third of the way up and a second at the top. Characters were wooden blocks 9 cm × 9 cm with 'googly eyes'. For 10-month-olds, the climber was a yellow triangle; helper and hinderer were a red square and a blue circle (counterbalanced). For 6-month-olds, the climber was a blue circle; helper and hinderer were a yellow triangle and a red square.

The curtain was first raised and lowered three times with the climber at the base of the incline. Habituation trials then began. The climber climbed to the lower plateau, 'danced' (jiggled up and down) for 2 s, then attempted twice to reach the upper plateau, each time falling back to the lower plateau. On a third attempt, the climber was either pushed to the top by the helper, or pushed to the bottom by the hinderer. In helping events, the helper entered the display from the lower right, moved up the incline and bumped the climber twice, each time pushing him farther up until the climber reached the upper plateau. The climber subsequently 'danced' while the helper went downhill and exited; the climber then became stationary. In hindering events, the hinderer entered from the upper left, moved down the incline and bumped the climber twice, each time pushing him farther down. The climber then tumbled end-over-end to hill bottom and remained stationary, while the hinderer moved back up the hill and exited. Total event duration was 11 s.

A coder blind to the identities of helper and hinderer recorded infants' looking to the stationary climber. Infants' looking was measured from helper/hinderer's exit until end-of-trial, reached when (1) the infant looked away continuously for 2 s, or (2) 60 s elapsed. The curtain dropped to occlude the display between trials. Infants saw helping and hindering trials in alternating sequence until (1) the summed looking times on three consecutive trials after the first three dropped to less than half the summed looking on the first 3 trials, or (2) 14 trials had elapsed. Both age groups habituated in an average of 9 trials.

Infants were then given choice and looking time test measures. Choice: the coder presented the infant with the helper and the hinderer 25 cm apart on a white board, and asked "would you like to pick a toy?" Infants' choice was defined as the character touched first, as judged by the (blind) coder, with the constraint that the infant had to be looking at the toy during or immediately preceding the touch (to exclude possibly accidental touches during board exploration, and so on). All infants in experiment 1 made identifiable choices. Looking time: the test display contained a shallow symmetrical test hill (122 cm × 66 cm, rising 14 cm from lowest to highest point). The climber sat at the top centre of the hill, with the helper and hinderer resting at the bottom left and right of the hill. The climber moved back and forth (10 cm each way) along the crest of the hill twice, then danced (2 s) at the top centre. Test trials then commenced. The climber partially approached, retreated, then fully approached to rest next to the helper or hinderer (on alternating trials); infants' looks to the now-stationary characters were then recorded.

A second coder independently coded a random 25% of subjects, achieving 99% agreement with the first coder on looking time and 100% on choice for both age groups.

The following were counterbalanced across subjects in each age group: (1) colour/shape of helper and hinderer; (2) order of helping and hindering habituation events; (3) order of choice and looking time measures; (4) positions of

helper and hinderer in choice and in looking time trials; and (5) order of 'approach-helper' and 'approach-hinderer' looking time trials.

Experiment 2: pushing inanimate object uphill versus downhill. Subjects were 12 10-month-olds (eight girls; mean age 10 months, 2 days; range 9 months, 14 days to 10 months, 22 days) and 12 6-month-olds (6 girls; mean age 6 months, 1 day; range 5 months, 15 days to 6 months, 17 days). Two additional 6-month-olds were excluded from the final sample owing to procedure error (1) and fussiness (1).

Stimuli were as in experiment 1 except that to create the 'object' stimulus, the eyes of the climber stimulus from experiment 1 were removed so that it appeared inanimate. Helper and hinderer stimuli from experiment 1 were used as pusher-up and pusher-down stimulus characters. Events were as in experiment 1 except that the object, unlike the climber in experiment 1, underwent no self-generated motion, and the pusher-up and pusher-down smoothly pushed the object all the way up (down) the incline from its starting location at bottom (top). Event duration was 11 s. In addition to being instructed not to interfere with their infants, parents were instructed to keep their eyes closed during the choice measure.

A second coder independently coded a random 25% of subjects of each age group; the two coders reached 100% agreement on choice, for both the 6- and 10-month-olds.

The following were counterbalanced across infants within each age group: (1) colour/shape of pusher-up and pusher-down; (2) order of habituation events; and (3) position of pusher-up and pusher-down in choice.

Experiment 3: valenced (helper/hinderer) versus neutral character. Subjects were 16 10-month-olds (eight girls; mean age 9 months, 27 days; range 9 months, 12 days to 10 months, 14 days) and 16 6-month-olds (eight girls; mean age 5 months, 28 days; range 5 months, 3 days to 6 months, 16 days). Eight additional infants (one 10-month-old) were excluded from the final sample owing to a procedure error (three), fussiness (two) and failure to make an identifiable choice (three). Half the subjects in each age group saw helping and neutral events, half saw hindering and neutral events.

For 10-month-olds, the climber was a blue circle; valenced and neutral characters were a red square and a yellow triangle. For 6-month-olds, the climber was a red circle; valenced and neutral characters were a yellow triangle and a blue square.

During habituation, half the infants (randomly chosen) in each age group saw helping and neutral events on alternate trials, half saw hindering and neutral events. Helping and hindering events were as in experiment 1. During neutral trials in the helping/neutral condition, the climber 'danced' for 2 s at the bottom of the incline (lower right), then sat motionless. The neutral character then entered from the lower right, bypassed the climber, and traced the same path as the helper, performing the same motions but not interacting with the climber. During neutral trials in the hindering/neutral condition, the climber 'danced' for 2 s at the lower right then sat motionless. The neutral character then entered from the upper left and traced the same path as the hinderer, performing the same motions without interacting with the climber.

Infants then received choice and looking time measures. In the choice measure, infants were presented with the valenced (helping or hindering) and neutral characters. In the looking time measure, the climber approached valenced and neutral characters on alternating trials.

A second coder independently coded a random 25% of subjects of each age group; the two coders reached 98% agreement on looking time and 100% agreement on choice, for both the 6- and 10-month-olds.

The following were counterbalanced across infants within each age group and condition: (1) colour/shape of valenced and neutral characters; (2) order of habituation events; (3) order of choice and looking time measures; (4) position of neutral character in choice and in looking time trials; and (5) order of 'approach-valenced' and 'approach-neutral' looking time trials.

LETTERS

Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite

Falk Warnecke^{1*}, Peter Luginbühl^{2*}, Natalia Ivanova¹, Majid Ghassemian², Toby H. Richardson^{2†}, Justin T. Stege², Michelle Cayouette², Alice C. McHardy^{3†}, Gordana Djordjevic², Nahla Aboushadi², Rotem Sorek¹, Susannah G. Tringe¹, Mircea Podar⁴, Hector Garcia Martin¹, Victor Kunin¹, Daniel Dalevi¹, Julita Madejska¹, Edward Kirton¹, Darren Platt¹, Ernest Szeto¹, Asaf Salamov¹, Kerrie Barry¹, Natalia Mikhailova¹, Nikos C. Kyrpides¹, Eric G. Matson⁵, Elizabeth A. Ottesen⁶, Xinning Zhang⁵, Myriam Hernández⁷, Catalina Murillo⁷, Luis G. Acosta⁷, Isidore Rigoutsos³, Giselle Tamayo⁷, Brian D. Green², Cathy Chang², Edward M. Rubin¹, Eric J. Mathur^{2†}, Dan E. Robertson², Philip Hugenholtz¹ & Jared R. Leadbetter^{5*}

From the standpoints of both basic research and biotechnology, there is considerable interest in reaching a clearer understanding of the diversity of biological mechanisms employed during lignocellulose degradation. Globally, termites are an extremely successful group of wood-degrading organisms¹ and are therefore important both for their roles in carbon turnover in the environment and as potential sources of biochemical catalysts for efforts aimed at converting wood into biofuels. Only recently have data supported any direct role for the symbiotic bacteria in the gut of the termite in cellulose and xylan hydrolysis². Here we use a metagenomic analysis of the bacterial community resident in the hindgut paunch of a wood-feeding 'higher' *Nasutitermes* species (which do not contain cellulose-fermenting protozoa) to show the presence of a large, diverse set of bacterial genes for cellulose and xylan hydrolysis. Many of these genes were expressed *in vivo* or had cellulase activity *in vitro*, and further analyses implicate spirochete and fibrobacter species in gut lignocellulose degradation. New insights into other important symbiotic functions including H₂ metabolism, CO₂-reductive acetogenesis and N₂ fixation are also provided by this first system-wide gene analysis of a microbial community specialized towards plant lignocellulose degradation. Our results underscore how complex even a 1-μl environment can be.

All known termite species form obligate, nutritional mutualisms with diverse gut microbial species found nowhere else in nature³. Despite nearly a century of study, however, science still has only a meagre understanding of the exact roles of the host and symbiotic microbiota in the complex processes of lignocellulose degradation and conversion. Especially conspicuous is our poor understanding of the hindgut communities of wood-feeding 'higher' termites, the most species-rich and abundant of all termite lineages⁴. Higher termites do not contain hindgut flagellate protozoa, which have long been known to be sources of cellulases and hemicellulases in the 'lower' termites. The host tissue of all wood-feeding termites is known to be the source of one cellulase, a single-domain glycohydrolase family 9 enzyme that is secreted and active in the anterior compartments of the gut tract⁵. Only in recent years has research provided support for a role of termite gut bacteria in the production

of relevant hydrolytic enzymes. That evidence includes the observed tight attachment of bacteria to wood particles, the antibacterial sensitivity of particle-bound cellulase activity², and the discovery of a gene encoding a novel endoxylanase (glycohydrolase family 11) from bacterial DNA harvested from the gut tract of a *Nasutitermes* species⁶. Here, in an effort to learn about gene-centred details relevant to the diverse roles of bacterial symbionts in these successful wood-degrading insects, we initiated a metagenomic analysis of a wood-feeding 'higher' termite hindgut community, performed a proteomic analysis with clarified gut fluid from the same sample, and examined a set of candidate enzymes identified during the course of the study for demonstrable cellulase activity.

A nest of an arboreal species closely related to *Nasutitermes ephratae* and *N. corniger* (Supplementary Fig. 1) was collected near Guápiles, Costa Rica. From worker specimens, luminal contents were sampled specifically from the largest hindgut compartment, the microbe-dense, microlitre-sized region alternatively known as the paunch or the third proctodeal segment (P3; Fig. 1a). In the interest of interpretive clarity, we specifically excluded sampling from and analysis of the microbiota attached to the P3 epithelium and the other distinct microbial communities associated with the other hindgut compartments.

Total community DNA from pooled P3 luminal contents was purified, cloned and sequenced. About 71 million base pairs of Sanger sequence data were generated and assembled. The assembly was highly fragmented, with the largest contiguous fragment being only about 15 kilobases (kb) long, attesting to a complex community that limited the analysis to being largely centred on genes and gene modules. For a better association of phylogenetic markers with key functional genes, 15 fosmids were selected for further analysis after an initial end-sequencing screen. The fosmid data were also used in the training of the sequence composition-based classifier, PhyloPythia⁷, resulting in the classification of 9% of all contigs beyond the level of phylum. The high gene-coding density and the fact that 0.03% of contigs were classified as Arthropoda suggest minimal host DNA sequence in the data. Genes associated with activities expected to be present among epithelial bacterial communities, such as aerobic

¹DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598, USA. ²Veregen Corporation (formerly Diversa), 4955 Directors Place, San Diego, California 92121, USA. ³IBM Thomas J. Watson Research Center, PO Box 218, Yorktown Heights, New York 10598, USA. ⁴Oak Ridge National Laboratory, Biosciences Division, Oak Ridge, Tennessee 37831-6026, USA. ⁵Department of Environmental Science and Engineering, ⁶Division of Biology, Mailcode 138-78, California Institute of Technology, Pasadena, California 91125, USA. ⁷INBio, Instituto Nacional de Biodiversidad, Apdo. Postal 22-3100 Santo Domingo de Heredia, Costa Rica. [†]Present addresses: Synthetic Genomics, Inc., 11149 North Torrey Pines Road, Suite 100, La Jolla, California 92037, USA (T.H.R., E.J.M.); Max Planck Institute for Computer Science, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany (A.C.M.).

*These authors contributed equally to this work.

respiratory metabolisms⁸, were rare or absent. From this we conclude that the source of the nucleic acids and proteins analysed in this study is the P3 luminal microbiota.

A PCR-based survey was used to assess the community structure of the P3 lumen. No amplification of 16S rRNA genes was observed with the use of Archaea-specific primers. An analysis of about 1,750 bacterial 16S rRNA gene sequences amplified from the community DNA revealed a broad diversity of bacteria representing 12 phyla and 216 phylotypes (99% sequence identity threshold) (Fig. 1b). Non-parametric diversity estimates indicated that about 80% of the total species diversity had been sampled (Supplementary Table 1 and

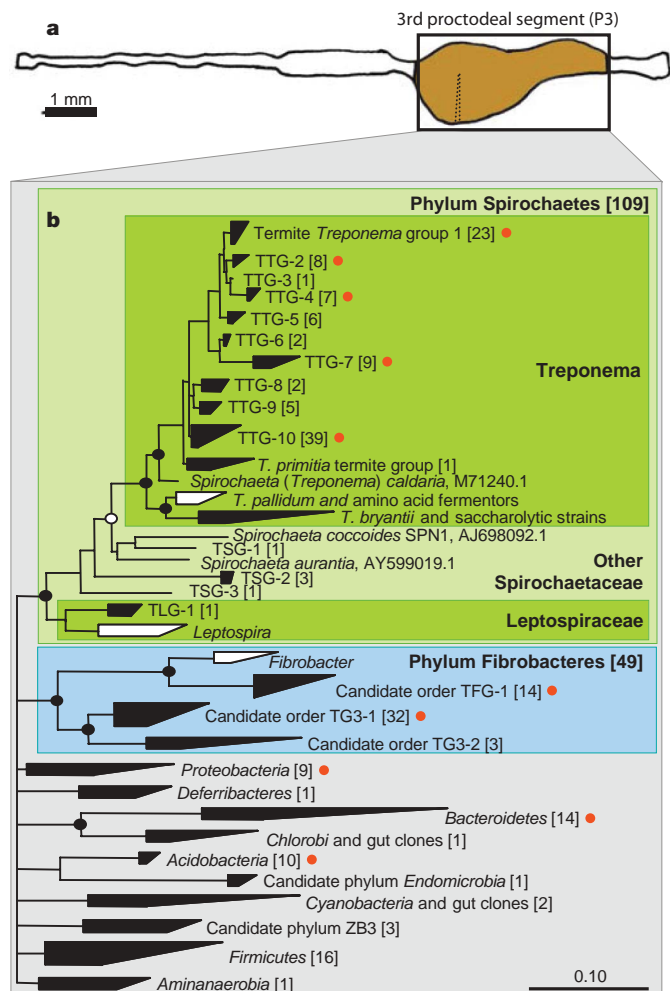


Figure 1 | Source and composition of the bacterial community analysed. **a**, Diagram of the gut tract extracted from the *Nasutitermes* host insect. The anterior of the P3 paunch was incised as indicated (dotted line) with a syringe needle. The luminal contents from about 165 specimens were collected and pooled for nucleic acid and protein analyses. **b**, Phylogenetic diversity of the P3 luminal microbiota. From a PCR-based inventory and from the metagenome libraries, 1,703 almost full-length and 41 full-to-partial 16S rRNA gene sequences were used in a maximum-likelihood analysis (RAxML). The number of distinct P3 luminal community phylotypes contained within each grouping is given in brackets (a total of 216 operational taxonomical units sharing at least 99% sequence identity); black shading denotes that at least one of these phylotypes was represented within the PCR inventory or metagenome libraries (red dots). White shading denotes reference groups having no representation in this or other public inventories derived from termite gut. The phylogram was constructed from 1,289 unambiguously aligned and filtered nucleotide positions. Branching pattern confidence is visualized as follows: no circle, more than 50%; open circle, more than 70%; filled circle, more than 90%. Scale bar indicates 10% estimated sequence divergence. See Supplementary Tables 1, 2 and 11 and Supplementary Figs 2–9 for accession numbers and detailed phylogenetic analyses.

Supplementary Fig. 2). The ten most frequently recovered phylotypes comprised 47% of the collection. The 24 most frequently recovered phylotypes in the PCR inventory belonged to the genus *Treponema* or the phylum Fibrobacteres (Fig. 1b). It was expected that members of these two phyla would constitute the majority of the source community genomic DNA analysed. Inspection of the metagenomic data for conserved phylogenetic marker genes and PhyloPythia binning of the data set supported this prediction. *Treponemes* dominated, contributing 68% of marker genes in the metagenomic DNA (Supplementary Table 2). Fibrobacteres contributed 13% of the identifiable fragments in the metagenomic data set.

Lignocellulose degradation requires a broad array of enzymes and associated proteins⁹. Consistent with past studies¹⁰, analysis of the *Nasutitermes* P3 luminal community metagenome revealed no evidence for lignin degradation. In contrast, a conservative analysis, detecting complete domains by global alignment (see Supplementary Information), identified many genes and gene modules homologous with more than 700 glycoside hydrolase (GH) catalytic domains corresponding to 45 different CAZy families (carbohydrate-active enzymes; <http://www.cazy.org>), including a rich diversity of putative cellulases and hemicellulases (Table 1 and Supplementary Tables 3–6).

More than 100 gene modules relevant to cellulose hydrolysis were identified in the data set, corresponding to the catalytic domains of GH5 cellulases, GH94 cellobiose or cellodextrin phosphorylases, GH51 endoglucanase/arabinofuranosidases, and a smaller number of GH8, GH9, GH44, GH45 and GH74 endoglucanases. In contrast, gene modules corresponding to the catalytic domains of GH6 and GH48 family endoglucanases and cellobiohydrolases, key components of several well-studied microbial cellulase systems, were absent.

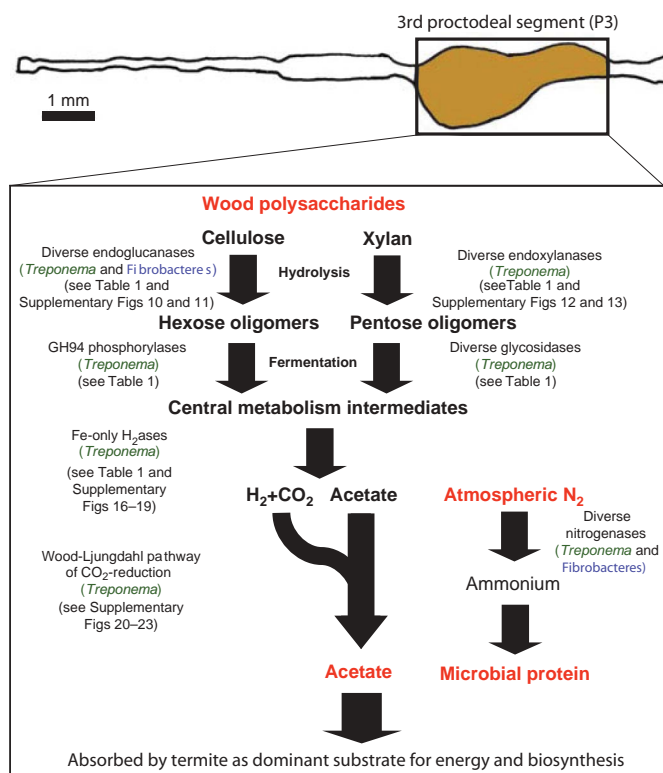


Figure 2 | Model of nutritional symbiosis-relevant metabolism by *Nasutitermes* P3 luminal bacteria. Wood is triturated by the insect's mandibles into small particles and predigested by poorly studied upstream processes before transit into the P3 compartment. The P3 lumen is dominated by diverse species of *Treponema* (Spirochaetes) and Fibrobacteres. There was no evidence for methanogenesis or lignin degradation in the metagenomic data set.

Table 1 | Glycoside hydrolases and carbohydrate-binding modules

CAZy family*	Pfam HMM name†	Known activities‡	Termite gut community§	Source organisms	Proteomics¶	Activity#	Signal peptides*
Glycoside hydrolase catalytic domains**††							
GH1	Glyco_hydro_1	β -Glucosidase, β -galactosidase, β -mannosidase, others	22	5 trep	1		
GH2	Glyco_hydro_2_C	β -Galactosidase, β -mannosidase, others	23	4 trep			1
GH3	Glyco_hydro_3	β -1,4-Glucosidase, β -1,4-xylosidase, β -1,3-glucosidase, α -L-arabinofuranosidase, others	69	14 trep	1		22
GH4	Glyco_hydro_4	α -Glucosidase, α -galactosidase, α -glucuronidase, others	14	3 trep	2		2
GH5	Cellulase	Cellulase, β -1,4-endoglucanase, β -1,3-glucosidase, β -1,4-endoxylanase, β -1,4-endomannanase, others	56	6 trep	2	26 of 30	14
GH8	Glyco_hydro_8	Cellulase, β -1,3-glucosidase, β -1,4-endoxylanase, β -1,4-endomannanase, others	5				3
GH9	Glyco_hydro_9	Endoglucanase, cellobiohydrolase, β -glucosidase	9	2 fibr		7 of 12	3
GH10	Glyco_hydro_10	Xylanase, β -1,3-endoxylanase	46	11 trep	1		8
GH11	Glyco_hydro_11	Xylanase	14	4 trep			
GH13	Alpha-amylase	α -Amylase, catalytic domain, and related enzymes	48	6 trep	2		
GH16	Glyco_hydro_16	β -1,3(4)-Endoglucanase, others	1				
GH18	Glyco_hydro_18	Chitinase, endo- β -N-acetylglucosaminidase, non-catalytic proteins	17	4 trep			6
GH20	Glyco_hydro_20	β -Hexosaminidase, lacto-N-biosidase	15	1 trep			
GH23	SLT	G-type lysozyme, peptidoglycan lytic transglycosylase	52	11 trep			9
GH25	Glyco_hydro_25	Lysozyme	1	1 trep			
GH26	Glyco_hydro_26	β -1,3-Xylanase, mannanase	15	3 trep			3
GH27	Melibiose	α -Galactosidase, α -N-acetylglactosaminidase, isomaltotransferase	4				
GH28	Glyco_hydro_28	Polygalacturonase, rhamnogalacturonase, others	6	1 trep			
GH31	Glyco_hydro_31	α -Glucosidase, α -xylosidase, others	26	7 trep			1
GH35	Glyco_hydro_35	β -Galactosidase	3	1 trep			
GH36	No Pfam§	α -Galactosidase, α -N-acetylglactosaminidase	5				
GH38	Glyco_hydro_38	α -Mannosidase	11	3 trep			1
GH39	Glyco_hydro_39	β -Xylosidase, α -L-iduronidase	3	2 trep			1
GH42	Glyco_hydro_42	β -Galactosidase	24	4 trep	1		
GH43	Glyco_hydro_43	Xylanase, β -xylosidase, α -L-arabinofuranosidase, arabinanase, others	16	6 trep			1
GH44	No Pfam§	Endoglucanase, xyloglucanase	6	2 trep			
GH45	Glyco_hydro_45	Endoglucanase (mainly eukaryotic, 2 bacterial)	4			2 of 2	2
GH51	No Pfam§	Endoglucanase, α -L-arabinofuranosidase	18	1 arch			
GH52	Glyco_hydro_52	β -Xylosidase	3				
GH53	Glyco_hydro_53	β -1,4-Endogalactanase	12	2 trep			
GH57	Glyco_hydro_57	α -Amylase, 4- α -glucanotransferase, α -galactosidase, others	17	6 trep			
GH58	No Pfam§	Endo-N-acetylneuraminidase or endo-sialidase	1				
GH65	Glyco_hydro_65m	Trehalase, maltose phosphorylase, trehalose phosphorylase	6				
GH67	Glyco_hydro_67M	α -Glucuronidase, others	10	4 trep			
GH74	No Pfam§	Endoglucanase, cellobiohydrolase, xyloglucanase	7				2
GH77	Glyco_hydro_77	4- α -Glucanotransferase, amylomaltase	14	4 trep			
GH88	Glyco_hydro_88	D-4,5 Unsaturated β -glucuronyl hydrolase	9	3 trep	1		
GH91	No Pfam§	Inulin fructotransferase	1				
GH92	Glyco_hydro_92	α -1,2-Mannosidase	2				
GH94	No Pfam§	Cellobiose phosphorylase, chitobiose phosphorylase, cellodextrin phosphorylase	68				
GH95	No Pfam§	α -L-Fucosidase	12				
GH98	Glyco_hydro_98M	Endo- β -galactosidase	1				
GH103	TIGR: MltB	Peptidoglycan lytic transglycosylase	3				
GH106	No Pfam§	α -L-Rhamnosidase	2				
GH109	No Pfam§	α -N-Acetylglactosaminidase	3				

A similar paucity of GH6 and GH48 modules has been observed in the genomes of other cellulose-degrading bacteria^{11,12}. A phylogenetic analysis of the GH5 cellulase diversity revealed that most constituted nine unique subclusters (Supplementary Fig. 10), none of which affiliated with *Clostridium* cellulosome GH5 cellulases. Functional genomic screens confirm that at least one example from each of five major, novel GH5 clades showed activity on acid-solubilized and microcrystalline cellulose (Table 1). At least 14 proteins with a GH5 module have an identifiable signal peptide, and metaproteomic analysis suggests that at least several are secreted into the P3 luminal fluid (Table 1). The metagenomic data suggest that the final metabolism of the oligosaccharides into simple sugars proceeds through the activity of GH3 glucosidase modules (at least 23 have an identifiable signal peptide), or through phosphorytic reactions catalysed by a diversity of GH94 cellobiose and cellodextrin phosphorylases after substrate transport into the bacterial cytoplasm.

About 100 gene modules corresponding to the catalytic domains of GH10, GH11, GH26 and GH43 hemicellulases were also identified (Table 1 and Supplementary Figs 12 and 13). The metaproteomic analysis revealed that at least one of the diverse GH10 endoxylanases

is expressed in the clarified P3 luminal fluid (Table 1); eight encode identifiable signal peptides. None of the xylanase modules have yet been examined for xylanase activity, but several cluster phylogenetically with a previously identified GH11 xylanase (Supplementary Fig. 13). Catalytic modules corresponding to several potential pentosidase and hemicellulose-debranching enzyme domains were also identified. In addition to the large number of GH3 modules noted above, many modules corresponding to the catalytic domains of GH1, GH2, GH4, GH67 and GH95 pentosidases were identified. Additionally, between 4 and 34 gene modules corresponding to the catalytic domains of carbohydrate esterase families 2, 4 and 6 acetyl xylan esterases were identified. Only a few gene modules for catalytic domains of polysaccharide lyases (PL) were observed (five PL1 and five PL11). On the basis of marker gene and composition-based binning analyses, we predict that most of the retrieved glycoside hydrolase gene modules are encoded by treponemes (Fig. 2 and Table 1). A notable exception to this is that the catalytic modules of GH45 and GH9 seemed to be encoded solely by fibrobacters.

Fewer putative polysaccharide-binding domains (carbohydrate-binding modules; CBMs) representing only five CAZy families (<http://www.cazy.org>) were identified in the data set. Modules of

Table 1 | Continued

CAZy family*	Pfam HMM name†	Known activities‡	Termite gut community§	Source organisms	Proteomics¶	Activity#	Signal peptides*
Carbohydrate-binding domains‡‡§§							
CBM4	CBM_4_9	Amorphous cellulose-, xylan- and glucan-binding domain	5		1		1
CBM6	CBM_6	Amorphous cellulose- and xylan-binding domain	13	2 trep			1
CBM11	CBM_11	Glucan-binding domain	3				
CBM30	No Pfam§	Cellulose-binding domain	1				
CBM32	F5_F8_type_C	Galactose- and lactose-binding domain	4	1 fibr			
CBM36	No Pfam§	Xylan-binding domain	2				
CBM37	No Pfam§	Broad binding specificity	1				
Other domains often associated with GH catalytic domains							
	Alpha-L-AF_C	α-L-Arabinofuranosidase, C-terminal domain; associated with GH51	10	1 trep			
	Alpha-mann_mid	Middle domain; associated with GH38	15	3 trep			
	Big_2	Bacterial Ig-like domain, group 2	140	7 trep			31
	Big_3	Bacterial Ig-like domain, group 3	22				
	Bgal_small_N	β-Galactosidase, small chain; associated with GH2	5				
	CBM_X	Associated with GH94	44	8 trep	8		
	CelD_N	N-terminal Ig-like domain of cellulase; associated with GH9	5	1 fibr			4
	fn3	Fibronectin type III domain	45	4 trep			3
	GDE_C	Amylo-α-1,6-glucosidase	1				
	Glyco_hydro_2	Immunoglobulin-like β-sandwich domain	4				
	Glyco_hydro_2_N	Sugar-binding domain	26	5 trep, 2 fibr			
	Glyco_hydro_3_C	C-terminal domain (glycan-binding?)	42	8 trep, 1 fibr	3		
	Glyco_hydro_38C	C-terminal domain	12	2 trep			
	Glyco_hydro_42C	C-terminal domain	8	1 trep			
	Glyco_hydro_42M	Trimerization domain	20	4 trep			
	Glyco_hydro_65C	C-terminal domain	3				
	Glyco_hydro_65N	N-terminal domain	1				
	Glyco_hydro_67C	C-terminal domain	11	2 trep			
	Glyco_hydro_67N	N-terminal domain	2				
	Glyco_transf_36	Associated with GH94	47	12 trep	8		
	GT36_AF	Associated with GH94	45	13 trep	8		
	He_PIG	Putative Ig-like domain	11	1 trep			
	Isoamylase_N	Isoamylase N-terminal domain, associated with GH13	15	2 trep			1

For a complete inventory and genome comparisons, see Supplementary Tables 3–6.

* CAZy: carbohydrate-active enzymes.

† Pfam, <http://www.sanger.ac.uk/Software/Pfam/>; HMM, hidden Markov model.

‡ <http://www.CAZy.org>.

§ Number of Pfam HMMs in the metagenomic data set with e-values smaller than 10^{-4} . For modules with no available Pfam HMM, this corresponds to the number of BLAST hits with e-values smaller than 10^{-6} . See Supplementary Information for representative 'non-Pfam' modules used in BLAST searches.

|| As determined by sequence composition-based classification using Phylopythia; see the text and Methods for details. (trep, treponeme; fibr, fibrobacter; arch, archaeal)

¶ Number of modules expressed as protein *in situ*, as demonstrated by a metaproteomic analysis of the clarified gut fluid.

Number of cloned modules having cellulase activity on PASC, out of the number examined for this activity.

*As determined using SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>); see Methods for details.

** CAZy GH families not represented in the data set: 6, 7, 12, 14, 15, 17, 19, 22, 24, 29, 30, 32, 33, 34, 37, 46, 47, 48, 49, 50, 54, 55, 56, 59, 61, 62, 63, 64, 66, 68, 70, 71, 72, 73, 75, 76, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 89, 90, 93, 96, 97, 99, 100, 101, 102, 107, 108 and 110.

†† CAZy GH families 21, 40, 41, 60, 69, 104 and 105 no longer exist, have been reassigned, or are recognized by the Pfam HMMs of other GH families.

‡‡ CAZy cellulose-binding domains (CBMs) not represented in the data set: 1, 2, 3, 5, 8, 10, 13, 14, 15, 17, 18, 19, 20, 21, 23, 24, 25, 26, 27, 29, 31, 33, 34, 35, 38, 39, 40, 41, 42, 43, 44, 45, 46, 48 and 49.

§§ CAZy CBMs 7, 9, 12, 16, 22, 28 and 47 either no longer exist or are recognized by the Pfam HMMs of other CBM families.

||| Several gene modules that are often found associated with glycoside hydrolases that were not represented in the data set include Cohesin, Dockerin_1, Alpha-amylase_C, CHB_HEX, CHB_HEX_C, ChiC, Chitinase_A_N, Glucodextran_B, Glucodextran_N, Glyco_hydro_20b, Glyco_hydro_32C, Glyco_hydro_98C and TIG.

undetermined function often found associated with various glycoside hydrolases were also identified (Table 1). *Clostridium*-like cellulosome domains, such as dockerins and cohesins, were not identified.

A total of 34 groups of genes of unknown function, termite orthologous groups, were found abundant in comparison with available metagenomes from soil¹³, seawater¹⁴ and human gut¹⁵ (Supplementary Table 10). The expression of several of these in clarified lumen fluid was verified by metaproteomic analysis (Supplementary Table 10). One uncharacterized, putative extracytoplasmic structural protein domain, TIGR02145, was strikingly overrepresented; it appeared 482 times in the termite gut metagenome (Supplementary Fig. 14). This module is found dozens of times in the genome of the rumen cellulolytic bacterium *Fibrobacter succinogenes* (<http://www.tigr.org/tdb/rumenomics>)¹⁶ but is otherwise restricted to only a few other microbes. It becomes tempting to speculate that this and several of the other classes of unknown domains highlighted in this study may be involved in cell-surface-associated enzymatic, binding and other functions relevant to lignocellulose conversion. However, this possibility has not yet been examined experimentally.

Free H₂ is generated as a product of the fermentation of cellulose and xylan by several protozoa from lower termites¹⁷, but its production rate, concentration and biological sources in protozoan-free, wood-feeding higher-termites are not yet known. Gene modules coding for 159 diverse iron-only hydrogenases¹⁸ were identified in

the metagenomic data set. Many were binned to the genus *Treponema* by PhyloPythia. By examining sequence similarity, domain structure and gene neighbourhoods, we predict that about half of these genes encode catalytic enzymes. Although some of the catalytic hydrogenases are closely related to known enzymes (Supplementary Fig. 15), several were inferred to have unusual spatial arrangements of the cysteines that coordinate the catalytic iron–sulphur clusters (Supplementary Figs 17 and 18). The rest seem to be involved in signal transduction and chemotaxis, as demonstrated by hydrogenases that include for example methyl-accepting chemotaxis domains (Supplementary Fig. 16).

The major electron sink reaction occurring during the fermentation of wood in termites is CO₂-reductive acetogenesis¹⁹. One-quarter of the gut acetate that ultimately serves as the major carbon and energy source for the insect host can be derived from CO₂ (ref. 20). Little is known about the diversity of CO₂-reducing acetogens in the gut tracts of higher, wood-feeding termites. Analysis of the data set identified 14–37 variants of all except one (formate dehydrogenase) of the known proteins associated with the Wood–Ljungdahl pathway²¹ of CO₂-reductive acetogenesis (Supplementary Fig. 19). Phylogenetic analysis of a marker gene of the pathway, that encoding formyl-tetrahydrofolate synthetase (FTHFS), revealed that most of the metagenome-derived genes clustered with orthologues from the genuine homoacetogen *Treponema primitia*

(Supplementary Fig. 21) and with other FTHFS genes that occur and are expressed in the guts of lower termites²². A key multi-enzyme complex of the pathway is carbon monoxide dehydrogenase (CODH) acetyl-CoA synthase²¹. Analysis of variants of hindgut *CooS* (the gene encoding the CODH catalytic subunit) revealed two distinct and novel phylogenetic clusters (Supplementary Fig. 22), both predicted to be encoded by treponemes by PhyloPythia. In the higher termite *Nasutitermes*, the symbiotic process of CO₂-reductive acetogenesis seems to be dominated by spirochetes, perhaps to the near or total exclusion of *Sporomusa termitida* or other spore-forming Firmicutes, historically the better-studied representatives of this physiology.

Genes encoding formate dehydrogenases (and related molybdopterin enzymes in general) and ion-translocating hydrogenases typically associated with the pathway were nearly absent from the data set. Formate is an obligatory pathway intermediate (Supplementary Figs 19 and 20), so the near absence of FDH remains unexplained. The scarcity of ion-translocating hydrogenases might be compensated for by the relative enrichment of RnfC-like ion-translocating NADH:ferredoxin oxidoreductases and ferredoxin-domain membrane proteins of unknown function (Supplementary Fig. 19).

Wood is depleted in nitrogen and is therefore a poor source of essential amino acids and protein. Bacterial N₂ fixation has been shown to affect the biology of Costa Rican *Nasutitermes* and other termite species²³, and a rich diversity of *nifH* genes has been inventoried in the hindguts of a *Nasutitermes* species from Japan²⁴. Twelve near-full-length *nifH* homologues were identified in this analysis of the P3 luminal community. Composition-based predictions did not successfully bin any of the P3 NifH homologues to known genera or phyla. Between 31 and 100 homologues of other nitrogenase components (NifD, NifK, NifE and NifN) were identified in the data set; at least one variant of each identified could be binned to the treponemes or to the fibrobacters, implicating species from both of these phyla in N₂ fixation in the gut (Fig. 2). The discrepancy between the contrasting recoveries of NifH and, for example, NifE homologues remains unexplained. The host's nitrogen waste, uric acid, has previously been shown to be fermented and recycled by *Streptococcus* and *Sebalidella* species resident in certain termite hindguts²⁵. However, there is as yet insufficient information on uric acid fermentation genes to be able to examine the data set for their presence.

Termite guts are characterized by steep physical, chemical and biological gradients²⁶. Human colonic and bovine ruminal microbial communities are dominated by non-motile microbes, but termite hindguts have long been noted as sources of highly motile bacteria²⁷. Motility in bacteria is often coupled with chemotactic behaviour. Indeed, about 1,500 genes related to chemotaxis and chemosensation were identified in the data set (Supplementary Fig. 23). The abundance of these genes in the termite gut suggests that chemotaxis may be relevant both during the colonization, by means of trophallaxis, of the microbe-free gut in young termites and after each moult, and also after regular peristaltic mixing and redistribution of the gut contents, thus maintaining compartmentalization and promoting niche selection.

This study illustrates how complex a 1-μl environment can be. The *Nasutitermes* species examined here and other termites may be considered to be rich reservoirs of bacterial enzymes relevant both to reaching a better understanding of the fundamental process of wood degradation, and to engineering novel schemes for the conversion of lignocellulose into biofuels and other microbial products of interest.

METHODS SUMMARY

Termites collected under permit near Guápiles, Costa Rica, were classified morphologically and by cytochrome oxidase gene sequence (Supplementary Fig. 1). Dissections were performed within 36 h of collection. Hemi-transverse incisions were made to the anterior P3 hindgut compartment. Luminal contents from about 165 worker specimens were pooled, diluted into buffered saline, and frozen immediately. Whole-genome shotgun libraries containing inserts were

prepared and sequenced. Data were assembled using Phrap; coding regions were predicted using fgenesb. Fosmids were nearly completely sequenced using pyrosequencing and traditional methods. Fragments 1 kb or larger were classified using PhyloPythia. Assembled data were incorporated into the Integrated Microbial Genomes with Microbiome Samples (IMG/M) system³² (<http://img.jgi.doe.gov/m>). For protein analysis, aliquots of the luminal contents in buffered saline were clarified by centrifugation. Protein from the soluble fraction was heat-treated, reduced, alkylated and digested. Digested samples were analysed by means of three-dimensional LC-MS/MS with XCalibur Rawfile Converter V, and searched against the termite metagenome database with SEQUEST. Searches for glycoside hydrolases and carbohydrate-binding modules were performed using HMMER *hmmsearch* with Pfam hidden Markov models (HMMs). GHs and CBMs were named using CAZy (carbohydrate-active enzymes) nomenclature. When a Pfam HMM for a given GH either did not exist or did not correspond to the catalytic module, BLAST searches against the data set were performed using regions of sequences listed at the CAZy website. Assays for activity were conducted with phosphoric-acid-swollen cellulose (PASC) and microcrystalline cellulose. 16S rRNA genes were amplified using universal primers, sequenced, and analysed with ARB and PHYLIP. High-quality sequence reads are deposited in the NCBI trace archive.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 July; accepted 18 September 2007.

1. Sugimoto, A., Bignell, D. E. & Macdonald, J. A. in *Termites: Evolution, Sociality, Symbioses, Ecology* (eds Abe, T., Bignell, D. E. & Higashi, M.) 409–435 (Kluwer Academic, Dordrecht, 2000).
2. Tokuda, G. & Watanabe, H. Hidden cellulases in termites: revision of an old hypothesis. *Biol. Lett.* **3**, 336–339 (2007).
3. Brune, A. in *The Prokaryotes* Vol. 1 (eds Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H. & Stackebrandt, E.) 439–474 (Springer, New York, 2006).
4. Kambhampati, S. & Eggleton, P. in *Termites: Evolution, Sociality, Symbioses, Ecology* (eds Abe, T., Bignell, D. E. & Higashi, M.) 1–24 (Kluwer Academic, Dordrecht, 2000).
5. Tokuda, G. *et al.* Major alteration of the expression site of endogenous cellulases in members of an apical termite lineage. *Mol. Ecol.* **13**, 3219–3228 (2004).
6. Brennan, Y. *et al.* Unusual microbial xylanases from insect guts. *Appl. Environ. Microbiol.* **70**, 3609–3617 (2004).
7. McHardy, A. C., Martin, H. G., Tsirigos, A., Hugenholtz, P. & Rigoutsos, I. Accurate phylogenetic classification of variable-length DNA fragments. *Nat. Methods* **4**, 63–72 (2007).
8. Brune, A., Emerson, D. & Breznak, J. A. The termite gut microflora as an oxygen sink: microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. *Appl. Environ. Microbiol.* **61**, 2681–2687 (1995).
9. Davies, G. J. & Henrissat, B. Structural enzymology of carbohydrate-active enzymes: implications for the post-genomic era. *Biochem. Soc. Trans.* **30**, 291–297 (2002).
10. Breznak, J. A. & Brune, A. Role of microorganisms in the digestion of lignocellulose in termites. *Annu. Rev. Entomol.* **39**, 453–487 (1994).
11. Taylor, L. E. II *et al.* Complete cellulase system in the marine bacterium *Saccharophagus degradans* strain 2-40T. *J. Bacteriol.* **188**, 3849–3861 (2006).
12. Xie, G. *et al.* Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*. *Appl. Environ. Microbiol.* **73**, 3536–3546 (2007).
13. Tringe, S. G. *et al.* Comparative metagenomics of microbial communities. *Science* **308**, 554–557 (2005).
14. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
15. Gill, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355–1359 (2006).
16. Qi, M. *et al.* Novel molecular features of the fibrolytic intestinal bacterium *Fibrobacter intestinalis* not shared with *Fibrobacter succinogenes* as determined by suppressive subtractive hybridization. *J. Bacteriol.* **187**, 3739–3751 (2005).
17. Odelson, D. A. & Breznak, J. A. Nutrition and growth characteristics of *Trichomitopsis termopsidis*, a cellulolytic protozoan from termites. *Appl. Environ. Microbiol.* **49**, 614–621 (1985).
18. Vignais, P. M. & Colbeau, A. Molecular biology of microbial hydrogenases. *Curr. Issues Mol. Biol.* **6**, 159–188 (2004).
19. Breznak, J. A. & Switzer, J. M. Acetate synthesis from H₂ plus CO₂ by termite gut microbes. *Appl. Environ. Microbiol.* **52**, 623–630 (1986).
20. Pester, M. & Brune, A. Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME J.* **1**, 551–565 (2007).
21. Ragsdale, S. W. The eastern and western branches of the Wood/Ljungdahl pathway: how the East and West were won. *Biofactors* **6**, 3–11 (1997).
22. Pester, M. & Brune, A. Expression profiles of *fhs* (FTHFS) genes support the hypothesis that spirochaetes dominate reductive acetogenesis in the hindgut of lower termites. *Environ. Microbiol.* **8**, 1261–1270 (2006).

23. Prestwich, G. D., Bentley, B. L. & Carpenter, E. J. Nitrogen sources for neotropical nasute termites: Fixation and selective foraging. *Oecologia* **46**, 397–401 (1980).
24. Ohkuma, M., Noda, S. & Kudo, T. Phylogenetic diversity of nitrogen fixation genes in the symbiotic microbial community in the gut of diverse termites. *Appl. Environ. Microbiol.* **65**, 4926–4934 (1999).
25. Potrikus, C. J. & Breznak, J. A. Anaerobic degradation of uric acid by gut bacteria of termites. *Appl. Environ. Microbiol.* **40**, 125–132 (1980).
26. Brune, A. & Friedrich, M. Microecology of the termite gut: structure and function on a microscale. *Curr. Opin. Microbiol.* **3**, 263–269 (2000).
27. Beckwith, T. & Light, S. The spirals within the termite gut for class use. *Science* **66**, 656–657 (1927).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank production and sequencing teams at Verenum and the Joint Genome Institute for their expertise, J. Mata for morphological identifications, L. Christoffersen for logistic and permitting support, and our laboratory colleagues for their comments during manuscript preparation. This research was supported in part by the National Science Foundation, the National Institutes of Health, Caltech, and the Lawrence Berkeley National Laboratory. The sequencing for the project was provided through the US Department of Energy (DOE) Community Sequencing Program (<http://www.jgi.doe.gov/CSP/index.html>). This work was performed, in part, under the auspices of the DOE Office of Science, Biological and Environmental Research Program, University of California, Lawrence Livermore National Laboratory, and Los Alamos National Laboratory.

Author Contributions F.W., P.H., E.J.M., D.E.R., E.M.R. and J.R.L. performed project planning, coordination, execution and facilitation. M.H., C.M., L.G.A. and G.T.

undertook field research planning, permits, logistics and station management. F.W., M.C., M.H., C.M., L.G.A. and J.R.L. conducted field collection and sample preparation. M.C., G.D., N.A., J.M. and C.C. performed nucleic acid purification and library construction. D.P. and K.B. carried out assemblies. A.S. conducted gene prediction and annotation. E.S. undertook data processing and loading into IMG/M. N.I. and N.C.K. performed metabolic reconstruction. A.C.M. and I.R. carried out binning. N.M. conducted fosmid annotation and manual curation. M.G., J.T.S. and B.D.G. performed proteomics and enzyme activities. F.W., P.L., V.K., D.D., E.K., E.G.M., E.A.O. and X.Z. carried out phylogenetic analyses. H.G.M. made accumulation curves, diversity estimates, statistical test for gene-centric analysis. P.L., T.H.R. and J.R.L. performed glycoside hydrolase bioinformatics. R.S. and S.G.T. constructed hypothetical gene families. N.I. and P.H. were responsible for hydrogenases. E.G.M., E.A.O., X.Z. and J.R.L. performed C1-pathway and N-metabolic reconstruction. M.P. carried out sensory transduction protein analysis. F.W., P.L., M.G., T.H.R., J.T.S., P.H., N.I., R.S., S.G.T., M.P. and J.R.L. undertook manuscript preparation.

Author Information This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under accession number ABDH00000000; this first version is ABDH01000000. The COII gene from the termite host is deposited under accession number EU236539, and 16S rRNA gene sequences are deposited under the accession numbers EF453758–EF455009 and EU024891–EU024927. The subcloned cellulase gene sequences are deposited under the accession numbers EF428062–EF428109. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.R.L. (jleadbetter@caltech.edu).

METHODS

Sample collection. Termite collection '290cost002' was made during the late morning of 24 May 2005 within the secondary forest of Bosque Lluvioso, an INBio private reserve near the town of Guápiles, Costa Rica (latitude 10° 11' 26.0" N, longitude 83° 51' 34.5" W). The nest was about 0.5 m in diameter and was attached to the trunk of an *Alchorneopsis floribunda* tree at a height of about 1 m. Specimens from this and other nests in the nearby area were observed to feed preferentially on dead limbs of *Cecropia obtusifolia*. At the time of collection, the temperature within the nest was 28 °C. The nest was cut into pieces with a machete and returned to InBio's laboratory in Heredia, where all dissections were performed within the first 36 h of collection. Specimens were identified morphologically as being most similar to *Nasutitermes ephratae*.

DNA extraction. For bacterial DNA and luminal protein analysis, gut tracts of ice-chilled, robust worker specimens were extracted using clean, sharp fine-tipped forceps, and arrayed on a piece of clean Parafilm. A hemi-transverse incision of the P3 hindgut compartment was made with the tip of a sterile 23-gauge needle attached to a 1-ml tuberculin syringe. Immediately thereafter, 7 µl of an anoxic buffered saline solution (BSS; 10.8 mM K₂HPO₄, 6.9 mM KH₂PO₄, 21.5 mM KCl, 24.5 mM NaCl, 0.5 mM CaCl₂, 10.0 mM NaHCO₃, 5.3 mM MgCl₂ pH 7.2) was placed over the incision; the BSS had been filtered through a 0.2-µm pore-size filter into sterile, crimp-sealed tubes containing an atmosphere of 100% N₂. Tubes of BSS were vacuum-purged for six cycles with sterile 100% N₂ to lower the dissolved O₂ concentration. The tip of a sterile, 200-µl plastic pipette tip attached to a standard micropipetter was used to press the dense, tan-coloured luminal contents gently out of the P3 compartment. The contents were stirred into the BSS, aspirated into the pipette and pooled into a microcentrifuge tube maintained on ice, and frozen at -80 °C immediately thereafter. The pooled P3 luminal contents from 165 worker specimens were used in all subsequent analyses of P3 luminal contents. DNA was isolated from the pooled luminal contents as described previously²⁸. DNA was also extracted from heads of soldier termites collected from nest 290cost002 with phenol/chloroform to confirm the identity of the host species.

Metagenome processing: shotgun library preparation, sequencing and assembly. Two shotgun libraries from the P3 genomic DNA were prepared: a 2–4-kb insert library cloned into pBK-CMV and a roughly 32-kb insert fosmid library cloned into pCC1Fos (Epicentre Corp.)²⁹. Both libraries were sequenced with BigDye Terminators v3.1 and resolved with ABI PRISM 3730 (ABI) sequencers. 92,160 reads comprising 68.47 megabases (Mb) of phred Q20 sequence were generated from the small insert library, and 13,824 reads comprising 2.91 Mb of phred Q20 sequence were generated from the pCC1Fos library. The reads were base-called using phred version 0.990722.g (refs 30, 31), and megablasted against the NCBI nucleotide database and internal contaminant databases. Reads with significant hits on known contaminants and/or with fewer than 100 Q20 base pairs were eliminated from the data set. The remaining 95,324 reads were assembled using parallel phrap (<http://www.phrap.org>) compiled for the Sun operating system, version SPS 3.57. The resulting assembly consisted of 41,765 contigs covering 44.4 trimmed consensus megabase pairs. The longest contig was 14.7 kbp long and contained 72 reads. However, 25% of the reads remained as singlets. This draft quality assembly was used for all downstream analysis.

Full fosmid sequencing and assembly. Fifteen termite hindgut fosmids were chosen for full sequencing. The fosmids were grown overnight in Luria-Bertani (LB) broth, normalized for cell density and pooled together with 15 fosmids from different projects for DNA extraction using the Qiagen Genomic-tip 100/G kit in accordance with the manufacturer's instructions. Extracted DNA was directly sequenced with the 454 Life Sciences Genome Sequencer GS20 and about 30 million base pairs of about 100-bp sequence reads were generated from one run. In addition, the pooled fosmid DNA was hydrosheared, subcloned into pUC18 and end-sequenced, producing 7,680 Sanger reads. Sanger and GS20 sequences were co-assembled with the Forge assembler developed at JGI (D.P., unpublished observations). Seven complete and eight partly assembled termite hindgut fosmids, comprising 37 genomic fragments, were assembled and manually verified from the pooled data set.

Gene prediction. The termite hindgut metagenome and fosmid assemblies were annotated with the *ab initio* gene-calling program ggenesb (SoftBerry). Normally *ab initio* gene calling of isolate genomes trains on the data set being annotated; however, because metagenomes are multigenomic data sets, self-training generates low-quality results (data not shown). Instead, parameters were obtained from training on multiple bacterial isolate genomes to provide an 'average' bacterial coding preference and other sequence features such as Shine-Dalgarno sequences. The command string used was `bactg_ann.pl mixr_paths_newcog.list1 <sequence_file> 60`, where the sequence file is the fasta output of the assembled contigs (for example, in Phrap this is the fasta.screen.contigs file),

and 60 is the minimal length of predicted ORFs in base pairs, and `mixr_paths_newcog.list1` is a config file that contains information about used programs, databases, etc. This file contains reference to `gener.par`, which provides generalized 'bacterial' gene parameters. The annotation of the 37 fosmid contigs was curated manually.

Binning. Metagenomic fragments were binned (classified) using PhyloPythia, a phylogenetic classifier that uses a multi-class Support Vector machine (SVM) for the composition-based assignment of fragments at different taxonomic ranks⁷. Generic models for the ranks of domain, phylum and class were combined with sample-specific models for the clades *Treponema* and Fibrobacteres. The generic models represent all clades covered by three or more species at the corresponding ranks among the sequenced microbial isolates. The sample-specific models include classes for the dominant sample populations of *Treponema* and Fibrobacteres, as well as a class 'Other'.

The sample-specific models were each trained on sequence data obtained from fully sequenced fosmids identified using phylogenetic marker genes (see Phylogenetic analysis of conserved single-copy protein-coding genes, below) and in the case of the *Treponema*-specific model, also from fragments of two sequenced *Treponema* genomes (*T. denticola* ATCC 34505 and *T. pallidum* subsp. *pallidum* str. Nichols). Five sample-specific SVMs were created by using fragments of lengths of 3, 5, 10, 15 and 50 kb. All input sequences were extended by their reverse complement before computation of the compositional feature vectors. The parameters *w* and *l* were both set to 5 for the sample-specific models.

Twelve of 15 sequenced fosmids assigned unambiguously through analysis of phylogenetic marker genes were initially used for the training of sample-specific models (four fosmids assigned to Fibrobacteres, corresponding to 94.5 kb of sequence, and eight fosmids assigned to *Treponema*, corresponding to 183.6 kb of sequence). Input fragments of a particular length were generated from the fosmids by using a sliding window with a step size of one-tenth of the generated fragment size (for example 5 kb for 50-kb fragments). For the class 'Other', fragments from 340 sequenced isolates were used, excluding those of the *Treponema* genomes. Composition-based characterization of the complete fosmid sequences with the combined classifier confirmed the marker-gene-based placement for 13 of the 15 fosmids. On the basis of this extended set (five Fibrobacteres fosmids totalling 115.2 kb of sequence and eight *Treponema* fosmids totalling 203 kb of sequence), the final sample-specific models were computed. The final classifier consisting of the sample-specific and generic clade models was then applied to assign all fragments more than 1 kb of the sample. In case of conflicting assignments, preference was given to assignments of the sample-specific models.

Loading of data into IMG/M. The annotated metagenomic and fosmid sequences (including PhyloPythia assignments) were loaded as independent data sets into IMG/M³² (<http://www.jgi.doe.gov/m>), a data-management and analysis platform for genomic and metagenomic data based on IMG³³. KEGG pathways present in the metagenome were automatically assigned on the basis of EC numbers in the annotation, and pathways not included in the KEGG database were inferred from comparative analysis with other genomes, on the basis of sequence similarity and gene cluster structure conservation. Conserved domain and motif searches were performed with RPS-BLAST, using an e-value threshold of 10⁻² without low-complexity masking³⁴. Overlaps on the query gene were removed as follows: the most significant hit for a query gene was taken first, then the next most significant hit that did not overlap with the first, and so on, until all overlaps had been removed. SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>) was used to identify signal peptides and the information was added to the records of individual genes in IMG/M³⁵. No significant differences in signal peptide calls were observed with the Gram-negative and Gram-positive models. Both were tried because of the multi-species nature of the data set.

Metaproteomic analysis: protein extraction and digest. Aliquots of the P3 luminal contents in buffered saline were centrifuged for 10 min (4 °C) at 13,000 r.p.m. on a benchtop microcentrifuge to remove the insoluble fraction of the fluid. Protein (500 µg) from the soluble fraction was heat-treated at 100 °C for 5 min in the presence of 0.5% RapiGest. The samples were reduced with 1 mM dithiothreitol and alkylated with 0.4 mg ml⁻¹ iodoacetamide. After digestion with endoproteinase Lys-C (1/200 of the protein), the protein samples were further digested twice with trypsin (1/100 of the protein). The second batch of trypsin was added 4 h after the addition of the first batch. The RapiGest was removed from the solution by the addition of HCl to a final concentration of 0.05 M. Samples were incubated at 37 °C for 60 min and centrifuged at 13,000 r.p.m. on a benchtop microcentrifuge to remove the insoluble fraction. The pH of the soluble fraction was adjusted to pH 2–3 with ammonium hydroxide.

Three-dimensional LC-MS/MS analysis. The digested termite protein samples were analysed by the three-dimensional LC-MS/MS system as described³⁶. The LTQ mass spectrometer was set to divide the full MS scan into three smaller

sections covering a total range of 400–2,000 *m/z*. Each of the smaller MS scans was followed by four MS/MS scans of the most intense ions from the preceding MS scan. The typical collision energy for collision-induced dissociation was set to 35% with a 30-ms activation time. Dynamic exclusion was enabled with a repeat count of 1 and a 3-min exclusion duration window.

Data analysis. The LC–MS/MS raw data were extracted with the XCalibur Rawfile Converter V 1.0.0a and then searched against the termite metagenome-predicted open reading frames using the SEQUEST program. The non-specific cleavage rule was designated during the SEQUEST search. Differential modifications of Met oxidation (+16) and Cys alkylation (+57) were allowed for the database search. The results were filtered using the same criteria set as described previously³⁶ to obtain the peptide identifications. In brief, all peptide identifications had a delta cross-correlation score of more than 0.1; peptides with a +1 charge state were fully tryptic with a cross correlation (X_{corr}) of more than 1.9; peptides with a +2 charge state were partly tryptic with $X_{\text{corr}} > 3.0$ or fully tryptic with X_{corr} ranging between 2.2 and 3.0; peptides with a +3 charge state were fully or partially tryptic with $X_{\text{corr}} > 3.5$. The Portfolio program was used for protein identification and summarizing the results. For each protein sequence in the database, the program assembled the peptide identifications that matched its substring. The protein identification was established if one or more peptide identifications were matched. The peptide identifications were allowed for multiple protein identifications. The number of distinct peptides and sequence coverage per protein identification were also calculated. Proteins identified by a single peptide are considered at a reduced level of confidence. These protein identifications by single peptide were included in all results.

Microbial community composition analysis and termite host identity: 16S rRNA gene PCR clone libraries. Two clone libraries were prepared from pooled PCR products amplified from the P3 genomic DNA using two different primer pairs broadly targeting the bacterial domain, namely 27F (5'-AGA GTT TGA TCC TGG CTC AG-3') and 1492R (5'-GGT TAC CTT GTT ACG ACT T-3'), and GM3 (5'-AGA GTT TGA TCM TGG C-3') and GM4 (5'-TAC CTT GTT ACG ACT T-3'). Attempts to amplify archaeal 16S rRNA genes were unsuccessful. PCR amplicons were cloned into the vector pCR4-TOPO (Invitrogen Corp.) and plated onto Carbenicillin-containing (150 µg ml⁻¹) LB agar plates, and five 384-well microtitre plates were picked. Clones were end-sequenced using vector primers, and the sequence reads were vector trimmed, assembled, quality-checked and chimera-checked using the genelib software package (E. Kirton, unpublished observations). Ten putative chimaeras were identified and excluded from the data set. A total of 1,703 near-complete bacterial 16S rRNA genes passed the quality and chimera filters in genelib and were used in the subsequent analyses.

Phylogenetic analysis of 16S rRNA gene sequences. The 1,703 sequences were aligned using the NAST aligner³⁷ and imported into an ARB database with the same alignment (<http://greengenes.lbl.gov/>; ref. 38). Forty-one partial and near complete 16S sequences were extracted from the P3 metagenomic data set using the NAST aligner and also imported into ARB⁴⁸. Sequences were initially assigned to phylogenetic groups using the ARB Parsimony insertion tool. *De novo* phylogenetic trees (Fig. 1 and Supplementary Figs 3 and 4) were constructed from masked ARB alignments (to remove ambiguously alignable positions) using RAXML³⁹. The bootstrapped tree was calculated using the following settings: rapid hill climb, general time-reversible model and optimization of site-specific evolutionary rates with 100 bootstrap resamplings while we used simulated annealing with the same settings but with the time limit of four times the average run time of the rapid hill-climb calculation for the snapshot tree. Snapshot and bootstrapped RAXML topologies were subsequently reimported in Newick format into ARB for visualization. The phylum-level trees (Supplementary Figs 5–9) were reconstructed using TREE-PUZZLE⁴⁰ in ARB.

Accumulation curves and diversity estimates. 16S rRNA gene sequences from the two PCR clone libraries were assigned to clusters (operational taxonomical units; OTUs) at 97%, 98%, 99% and 100% sequence identity thresholds using the DOTUR package⁴¹ applied to a distance matrix generated from an unmasked alignment in ARB (Supplementary Table 1). Accumulation curves were constructed by plotting the number of unique OTUs at a 99% identity threshold found in a given number of clones versus the number of clones (Supplementary Fig. 2). The clone order is arbitrary, so the accumulation curve was averaged over a 1,000 random permutations of the initial clone order. Accumulation curves were also plotted for subsets of the OTUs comprising the major phylogenetic lineages identified in the data set. Diversity estimates were calculated using four different methods involving parametric (exponential⁴² and two-parameter hyperbola^{42,43}) and non-parametric (bootstrap^{44–46} and jackknife^{44,45}) estimators (Supplementary Table 1). Neither the exponential nor the two-parameter hyperbola models offered good fits, so they were not considered further. The non-parametric estimates for these data produced more meaningful estimates (errors were much smaller than averages) of the total diversity (Supplementary

Table 1). On the basis of these estimates, 71–87% of the total OTU diversity had been sequenced.

Phylogenetic analysis of conserved single-copy protein-coding genes. Orthologues belonging to ten conserved single-copy protein-coding gene families were retrieved from the P3 metagenomic and isolate genome data sets using an export tool available in IMG/M³². Genes from the termite data set had to be more than 25% of their full-length orthologue in *Escherichia coli* to be included in the analysis. Multiple sequence alignments were created using Muscle⁴⁷ and phylogenetic trees were reconstructed using the ARB software package⁴⁸. If sequences clustered clearly with the available *Treponema* sequences (*T. pallidum* and *T. denticola*) or *Fibrobacter succinogenes*, they were assigned to these phylum-level groupings. If no clear relationship with these phyla could be established they were assigned to 'other' (see Supplementary Table 2). Note that these 'other' sequences could still belong to either the Fibrobacteres or Spirochaeta because the available reference genome sequences are related only distantly to the termite hindgut bacteria.

Identification of termite host species. The mitochondrial cytochrome oxidase II (COII) gene was PCR-amplified from genomic DNA extracted from termite soldier heads using the primers A-tLeu_mod (5'-CAG ATA AGT GCA TTG GAT TT-3') and B-tLys (5'-GTT TAA GAG ACC AGT ACT TG-3') and was sequenced without cloning. This gene is commonly used for the identification of termite species^{49,50}. Reference COII nucleotide sequences were extracted from the public databases and aligned to the termite sequences in ARB. Nucleotide and amino-acid-based trees were constructed using TREE-PUZZLE in ARB (Supplementary Fig. 1) and the topologies were compared.

Glycoside hydrolases and carbohydrate-binding modules: annotation. Database searches for GHs and CBMs were performed using HMMER hmsearch with Pfam HMMs. The Pfam_ls HMMs were used to find complete matches to the family by global alignment. All hits with *e*-values smaller than 10⁻⁴ were counted and their sequences were further analysed. GHs and CBMs were named in accordance with the CAZy nomenclature scheme^{9,51,52}. The data sets analysed comprised the 62-Mb assembly of the termite P3 hindgut metagenome, the human distal gut microbiome (subjects 7 and 8), the soil (Diversa silage) metagenome, the completed genomes of *Saccharophagus degradans* (*Microbulifer degradans*), *Cytophaga hutchinsonii* and *Thermobifida fusca*, and the draft genomes of *Fibrobacter succinogenes*, *Clostridium thermocellum* and *Caldicellulosiruptor saccharolyticus*. The closest related protein sequences for all metagenome sequences were identified with BLAST searches in GenBank and added to the data set. For several CAZy GH and CBM families, no Pfam HMM is available. In these cases, representative sequences were selected from the CAZy website, and the sequence region corresponding to the family of interest was determined. These sequence regions were then used in BLAST searches against the termite gut community metagenome to identify potential sequences belonging to these CAZy GH and CBM families. An *e*-value cutoff of 10⁻⁶ was used in these searches. The sequences of the domains used in these searches are listed in Supplementary Information.

Phylogenetic analysis. Sequence alignments were obtained using HMMER hmalign to the corresponding Pfam HMM. Regions that did not align to the Pfam HMMs were masked from the final alignments. Phylogenetic analysis was performed with the program PROML, which is part of the PHYLIP package, using a maximum-likelihood method with the Jones–Taylor–Thornton (JTT) amino-acid change model, equal rates, and one random sequence addition followed by global rearrangements.

Subcloning and expression of glycoside hydrolases and activity assay. Subclone sequences that were discovered in the termite gut community DNA library by hybridization or activity screening, or were selected from termite gut community metagenome (data mining) and retrieved in full length, were assayed for activity on PASC^{33,54}, and were added to the data set as well. Of the 33 active enzymes, 32 were identified through heterologous expression from the pSE420 vector in *E. coli*, the other (GenBank EF428075) was identified through expression in *Pichia pastoris*. Whole-cell lysates were prepared and tested for cellulase activity on a simple PASC substrate (50 mM sodium acetate pH 5 or 50 mM NaPO₄ pH 7, 0.75% PASC, 1:25 dilution of cell lysate, 37 °C overnight)⁵⁵. Activity was monitored by glucose oxidase/β-glucosidase enzyme couple reactions for cellobiose and glucose release or with the bicinchoninic acid assay (BCA) for reducing-sugar formation. Cellulose assays for glucose/cellobiose release employed the Invitrogen Amplex Red Glucose/Glucose Oxidase Assay Kit (A22189) or the BCA assay kit (Pierce 23221). Enzymes were considered positive for cellulase activity if they produced a signal in these assays greater than one standard deviation from the average activity observed by a vector-only control lysate.

Hypothetical gene families. For the analysis of overrepresented gene families of unknown function, the STRING database⁵⁶ was first used to assign COG (Clusters of Orthologous Groups) or NOG (Non-supervised Orthologous

Group) accession to each protein in the set. COG/NOG groups containing 50 or more members in the termite gut microbiota genes and belonging to functional categories 'S' ('function unknown') or 'R' ('general function prediction only') were further analysed. Proteins that were not assigned any COG/NOG were clustered using an all-versus-all blastp (default parameters) followed by mcl clustering as described (<http://micans.org/mcl/>). Representation of each of the above-described clusters of homologues was calculated in the soil¹³ and seawater¹⁴ metagenomes, and clusters of hypothetical proteins that showed at least threefold overrepresentation in the termite gut microflora sequences were considered as termite orthologous groups (TOGs; Supplementary Table 12).

The statistical significance of gene-family overrepresentation was determined by first assigning a relative representation to each orthologous group. The relative family (COG, NOG or TOG) representation r_{ij} ($i = 1, \dots, F$ is a unique number given to each COG, NOG and TOG, and $j = 1, \dots, 4$ is a number given to each of the environments) was calculated by normalizing the total number of ORFs of a family i in an environment j , C_{ij} , by the total number of ORFs in each environment:

$$p_{ij} = C_{ij} / \sum_{i=1}^F C_{ij}$$

and then normalizing to unity all environments:

$$r_{ij} = p_{ij} / \sum_{j=1}^4 p_{ij}$$

The relative family representation pinpoints families that are overrepresented in one data set with respect to the others: for example, a value of $r_{i1} = 0.8$ indicates that 80% of the normalized ORF counts of family i are found in the termite hindgut environment (first environment) and 20% are spread over the remaining environments.

Additionally, it would be desirable to have a measure of the robustness of the family representation value. For example, if for two families k and l the counts were $C_{k(j=1,\dots,4)} = [1\ 0\ 0\ 0]$ and $C_{l(j=1,\dots,4)} = [423\ 0\ 0\ 0]$, both family representation values r_{j1} and r_{k1} would equal 1 because there are no counts for other data sets. Nonetheless, the result for the second family would be more reliable because it involves a larger number of ORF counts. It is therefore less likely to have been a consequence of noise or finite sampling effects. We quantify this robustness by using a null model in which the possible effects due to finite sampling effects and noise are modelled in the simplest way: by taking 1% of randomly selected ORF counts and reassigning them to randomly chosen families and environments. This is not meant to be a comprehensive modelling of all sources of error, but rather using a basic null model that tests robustness of estimates by introducing a small amount of noise.

The addition of noise can be formally described as follows: if P_{ijk} denotes the presence of ORF k in family i and environment j ($P_{ijk} = 1$ if ORF k belongs to family i and is present in environment j ; $P_{ijk} = 0$ otherwise), for a randomly chosen k belonging to family i and present in environment j $P_{ijk} = 1 \rightarrow 0$ and for randomly chosen l and m $P_{ijk} = 0 \rightarrow 1$. This procedure was repeated a total of $N = 10,000$ times and the relative representation r_{ij}^s for each of these realizations ($s = 1, \dots, N$) was calculated by normalizing as explained above. The average

$$r_{ij}^{\text{av}} = \sum_{s=1}^N r_{ij}^s / N$$

and standard deviation

$$\sigma = \sqrt{\sum_{s=1}^N [(r_{ij}^{\text{av}} - r_{ij}^s)^2] / N}$$

were calculated as usual. The standard deviation represents an estimate of how robust the relative representation is with respect to noise, and the average is an estimate of the relative representation corrected by the addition of noise.

A P value was calculated to sort the families according to the following relevance criteria: high relative representation in the termite hindgut environment and low standard deviation. The P value for the relative representation of a family i in an environment j was calculated by adding together the number of environments for each family and generation that had a relative normalization equal to or higher than the average r_{ij}^{av} and a standard deviation equal to or smaller than the standard deviation σ_{ij} for that family and environment, divided by the total number of possible positive hits ($4NF$)

$$p_{ij} = \sum_{k=1, \dots, F, l=1, \dots, 4, s=1, \dots, N} Q_{kls} / 4NF$$

where $Q_{kls} = 1$ if both $r_{kl}^s \geq r_{ij}^{\text{av}}$ and $\sigma_{kl} \leq \sigma_{ij}$, and zero otherwise. The comparisons were done by first binning the values of r_{ij}^{av} and σ_{ij} in intervals of size 0.01 and 0.005 respectively. In this way, a P value of, say, 0.05 for a given family means that only 5% of the families are more relevant to our analysis (higher representation in the termite environment and lower standard deviation) than the considered family.

The 1% error level is, admittedly, arbitrary. The only condition it must fulfil is that it not be large enough to distort the initial data set. A selection of different levels of noise (for example 3% or 0.1%) changes the standard deviations linearly (while the noise is kept low enough) but does not change the P value or family ordering. The pfam representation and P values were calculated in the same way as ORF families.

Hydrogenases. Analysis of abundance of hydrogenase families with different cofactor specificity was performed using the tools provided in IMG/M³². Genes in the termite hindgut metagenome were assigned to COG clusters and Pfam families using reverse position-specific BLAST (RPS-BLAST⁵⁷) against the NCBI Conserved Domain Database⁵⁸ with a maximal e-value of 0.1 and a minimal percentage identity of 20%. Genes containing iron-only hydrogenase catalytic domain according to COG and Pfam classification (COG4624 and pfam02906) were clustered into families based on a minimal percentage of sequence identity of 70% for proteins within the family. Proteins within each family were aligned and consensus sequences were generated using Multalin⁵⁹, because in our experience this outperforms other tools when aligning fragmented sequences. Domain composition of each family was determined using InterProScan⁶⁰ of the consensus sequences. The catalytic domain of iron-only hydrogenases comprising proximal Fe-S clusters FS4A, FS4B and active-site H cluster⁶¹ were aligned using ClustalW⁶²; the ARB software package⁴⁸ was used for alignment editing, generation of consensus sequences for each hydrogenase family and tree calculation. Three-dimensional comparative modelling of iron-only hydrogenase families was performed using 3D-JIGSAW server⁶³ and was based on the structure of *Clostridium pasteurianum* iron-only hydrogenase (PDB: 1FEH⁶¹). Visualization of modelled structures was performed using the VMD package⁶⁴.

H₂-dependent reduction of CO₂ to acetate. For the analysis of CO₂-reductive acetogenesis summarized in Supplementary Fig. 20, Wood-Ljungdahl pathway-associated proteins were blasted against the termite metagenome using the tools provided in IMG/M. The GenBank accession numbers and species for each of the alleles employed in this search are listed here. Formate dehydrogenases and other molybdopterin enzymes: P07658 (*E. coli*) and ABC20599 (*Moorella thermoacetica*). Formyl-THF synthetase: AAP55207 (*Treponema primitia* ZAS-2). Methenyl-THF cyclohydrolase: AAP55206 (*Treponema primitia* ZAS-2). Methylene-THF dehydrogenase: YP_430368 (*Moorella thermoacetica*). Methylene-THF reductase: NP_662255 (*Chlorobium tepidum* TLS). Methyl transferase (THF to Fe-S-Co protein): AAA53548 (*Moorella thermoacetica*). Large-subunit Fe-S-Co protein: Q07340 (*Moorella thermoacetica*). Small-subunit Fe-S-Co protein: AAA23255 (*Moorella thermoacetica*). CoH-like ion translocating NiFe hydrogenase: YP_360647 (*Carboxydotherrmus hydrogenoformans*). RnfC-like ion translocating oxidoreductase: CAA51399 (*Rhodobacter capsulatus*). CODH/ACS CooS-subunit: P31896 (*Rhodospirillum rubrum*). CoC Ni-insertion protein: YP_430061 (*Moorella thermoacetica*). Acetyl-CoA synthase (AcsR): P27988 (*Moorella thermoacetica*). CODH/ACS-associated membrane protein with ferredoxin domain (COG3894): YP_430057 (*Moorella thermoacetica*). For the phylograms presented in Supplementary Figs 21, 22 and 24, proteins from each family were aligned using ClustalW⁶². The ARB software package⁴⁸ was used for alignment editing, generation of sequence masks and calculation of phylograms for each.

Dinitrogen fixation. For the analysis of nitrogen fixation associated genes, key proteins were blasted against the termite metagenome using the tools provided in IMG/M. The GenBank accession numbers for each of the alleles employed in this search are listed here. *Clostridium beijerinckii* genome information was used as the source of full-length *NifD*, *E*, *H*, *K* and *N* alleles: AAS91667, AAS91669, AAF77055, AAS91668 and AAS91670, respectively. Partial *NifH* sequences from the databases were also blasted against the metagenome. The GenBank accession numbers for each of these truncated alleles are listed here. From ref. 24: BAA28469, clone NTN2; BAA28480, clone NTN6; BAA28467, clone NTN18; BAA28470, clone NTN21; BAA28474, clone NTN30; BAA11781, clone TDY5. From ref. 65: AAK01229, *Treponema primitia* ZAS-2; AAK01231, *Treponema azotonutricium* ZAS-9; AAK01222, *Spirochaeta stenostrepta*.

Signal transduction. Distribution of genes encoding signal transduction (ST) proteins in the various metagenomic data sets and that of representative completed genomes was based on hits to COG functional categories, which include not only full-length genes but also separate domains. The search for ST COG categories was done in IMG/M using the default parameters. The abundance data were normalized using the Z-score function in the IMG

Metagenomics database, which takes into account the number of genes in the different data sets. The resulting COG categories hits were further organized into broad functional categories (for example, chemotaxis, other types of two-component system families, individual kinases, phosphatases, sensors and regulators genes or domains). The data were visualized graphically using the software Genesis⁶⁶ as a two-dimensional matrix in which the greyscale intensity is proportional to the abundance value.

28. Robertson, D. E. *et al.* Exploring nitrilase sequence space for enantioselective catalysis. *Appl. Environ. Microbiol.* **70**, 2429–2436 (2004).
29. Short, J. M., Fernandez, J. M., Sorge, J. A. & Huse, W. D. Lambda ZAP: a bacteriophage lambda expression vector with *in vivo* excision properties. *Nucleic Acids Res.* **16**, 7583–7600 (1988).
30. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
31. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
32. Markowitz, V. M. *et al.* An experimental metagenome data management and analysis system. *Bioinformatics* **22**, e359–e367 (2006).
33. Markowitz, V. M. *et al.* The integrated microbial genomes (IMG) system. *Nucleic Acids Res.* **34**, D344–D348 (2006).
34. Marchler-Bauer, A., Panchenko, A. R., Ariel, N. & Bryant, S. H. Comparison of sequence and structure alignments for protein domains. *Proteins* **48**, 439–446 (2002).
35. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**, 953–971, doi:10.1038/nprot.2007.131 (2007).
36. Wei, J. *et al.* Global proteome discovery using an online three-dimensional LC-MS/MS. *J. Proteome Res.* **4**, 801–808 (2005).
37. DeSantis, T. Z. Jr *et al.* NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res.* **34**, W394–W399 (2006).
38. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).
39. Stamatakis, A., Ludwig, T. & Meier, H. RAXML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**, 456–463 (2005).
40. Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504 (2002).
41. Schloss, P. D. & Handelsman, J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.* **71**, 1501–1506 (2005).
42. Colwell, R. K. & Coddington, J. A. Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. R. Soc. Lond. B* **345**, 101–118 (1994).
43. de Caprariis, P., Lindemann, R. & Collins, C. A method for determining optimum sample size in species diversity studies. *J. Int. Assoc. Math. Geol.* **8**, 575–581 (1976).
44. Smith, E. P. & van Belle, G. Nonparametric estimation of species richness. *Biometrics* **40**, 119–129 (1984).
45. Krebs, C. J. *Ecological Methodology* (Harper & Row, New York, 1989).
46. Efron, B. & Tibshirani, R. *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993).
47. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
48. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
49. Austin, J. W., Szalanski, A. L. & Cabrera, B. J. Phylogenetic analysis of the subterranean termite family *Rhinotermitidae* (Isoptera) by using the mitochondrial cytochrome oxidase II gene. *Ann. Entomol. Soc. Am.* **97**, 548–555 (2004).
50. Ohkuma, M. *et al.* Molecular phylogeny of Asian termites (Isoptera) of the families *Termitidae* and *Rhinotermitidae* based on mitochondrial COII sequences. *Mol. Phylogenet. Evol.* **31**, 701–710 (2004).
51. Henrissat, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316 (1991).
52. Davies, D. G. & Geesey, G. G. Regulation of the alginate biosynthesis gene *algC* in *Pseudomonas aeruginosa* during biofilm development in continuous culture. *Appl. Environ. Microbiol.* **61**, 860–867 (1995).
53. Johnston, D. in *Handbook of Food Enzymology* (eds Whitaker, J. R., Voragen, A. G. J. & Wong, D. W. S.) 761–769 (Marcel Dekker Inc., New York, 2002).
54. Stahlberg, J., Johansson, G. & Pettersson, G. *Trichoderma reesei* has no true exocellulase: all intact and truncated cellulases produce new reducing end groups on cellulose. *Biochim. Biophys. Acta* **1157**, 107–113 (1993).
55. Wood, T. Preparation of crystalline, amorphous and dyed cellulase substrate. *Methods Enzymol.* **160**, 19–25 (1988).
56. von Mering, C. *et al.* STRING 7—recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.* **35**, D358–D362 (2007).
57. Marchler-Bauer, A. & Bryant, S. H. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* **32**, W327–W331 (2004).
58. Marchler-Bauer, A. *et al.* CDD: a conserved domain database for protein classification. *Nucleic Acids Res.* **33**, D192–D196 (2005).
59. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881–10890 (1988).
60. Apweiler, R. *et al.* The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **29**, 37–40 (2001).
61. Peters, J. W., Lanzilotta, W. N., Lemon, B. J. & Seefeldt, L. C. X-ray crystal structure of the Fe-only hydrogenase (CpI) from *Clostridium pasteurianum* to 1.8 angstrom resolution. *Science* **282**, 1853–1858 (1998).
62. Chenna, R. *et al.* Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31**, 3497–3500 (2003).
63. Bates, P. A., Kelley, L. A., MacCallum, R. M. & Sternberg, M. J. E. Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins Struct. Funct. Genet.* **45**, 39–46 (2001).
64. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 27–28 (1996).
65. Lilburn, T. G. *et al.* Nitrogen fixation by symbiotic and free-living spirochetes. *Science* **292**, 2495–2498 (2001).
66. Sturn, A., Quackenbush, J. & Trajanoski, Z. Genesis: cluster analysis of microarray data. *Bioinformatics* **18**, 207–208 (2002).

LETTERS

The inhibitory cytokine IL-35 contributes to regulatory T-cell function

Lauren W. Collison¹, Creg J. Workman¹, Timothy T. Kuo³, Kelli Boyd², Yao Wang¹, Kate M. Vignali¹, Richard Cross¹, David Sehly⁴, Richard S. Blumberg³ & Dario A. A. Vignali¹

Regulatory T (T_{reg}) cells are a critical sub-population of $CD4^{+}$ T cells that are essential for maintaining self tolerance and preventing autoimmunity^{1,2}, for limiting chronic inflammatory diseases, such as asthma and inflammatory bowel disease^{3,4}, and for regulating homeostatic lymphocyte expansion⁵. However, they also suppress natural immune responses to parasites⁶ and viruses⁷ as well as anti-tumour immunity induced by therapeutic vaccines⁸. Although the manipulation of T_{reg} function is an important goal of immunotherapy, the molecules that mediate their suppressive activity remain largely unknown. Here we demonstrate that Epstein-Barr-virus-induced gene 3 (*Ebi3*, which encodes IL-27 β) and interleukin-12 alpha (*Il12a*, which encodes IL-12 α /p35) are highly expressed by mouse $Foxp3^{+}$ (forkhead box P3) T_{reg} cells but not by resting or activated effector $CD4^{+}$ T (T_{eff}) cells, and that an *Ebi3*–IL-12 α heterodimer is constitutively secreted by T_{reg} but not T_{eff} cells. Both *Ebi3* and *Il12a* messenger RNA are markedly upregulated in T_{reg} cells co-cultured with T_{eff} cells, thereby boosting *Ebi3* and IL-12 α production *in trans*. T_{reg} -cell restriction of this cytokine occurs because *Ebi3* is a downstream target of *Foxp3*, a transcription factor that is required for T_{reg} -cell development and function. *Ebi3*^{−/−} and *Il12a*^{−/−} T_{reg} cells have significantly reduced regulatory activity *in vitro* and fail to control homeostatic proliferation and to cure inflammatory bowel disease *in vivo*. Because these phenotypic characteristics are distinct from those of other IL-12 family members, this novel *Ebi3*–IL-12 α heterodimeric cytokine has been designated interleukin-35 (IL-35). Ectopic expression of IL-35 confers regulatory activity on naive T cells, whereas recombinant IL-35 suppresses T-cell proliferation. Taken together, these data identify IL-35 as a novel inhibitory cytokine that may be specifically produced by T_{reg} cells and is required for maximal suppressive activity.

In a functional genomics screen to identify novel T_{reg} -cell-restricted molecules, we identified *Ebi3* as being significantly over-represented in mouse T_{reg} ($CD4^{+}CD25^{+}$) versus naive T_{eff} ($CD4^{+}CD25^{-}$) cells. A recent study also found that *Ebi3* was upregulated in $Foxp3^{+}$ T_{reg} cells⁹. *Ebi3* is a member of the IL-12 heterodimeric cytokine family and is structurally similar to the IL-12 β chain (also known as p40, which is encoded by *Il12b*)¹⁰. *Ebi3* pairs with IL-27 α (also known as p28, which is encoded by *Il27*), which resembles IL-12 α , to yield IL-27 (see schematic in Supplementary Fig. 1)¹¹. *Ebi3* was first identified in B cells infected with Epstein-Barr virus¹⁰, and its expression is largely restricted to cells of the myeloid lineage¹¹. No overt autoimmunity or inflammatory disease has been reported in *Ebi3*^{−/−} mice¹². However, a role for IL-27 in the induction of inflammation has been demonstrated¹² so the suppressive activity of a negative regulatory protein in these mice may be negated, leaving open the possibility that *Ebi3* has both pro-inflammatory and anti-inflammatory roles.

Real-time quantitative polymerase chain reaction (PCR) analysis confirmed the restricted expression of *Ebi3* in peripheral $CD4^{+}CD45RB^{lo}CD25^{+}$ T_{reg} cells versus naive $CD4^{+}CD45RB^{hi}CD25^{-}$ T_{eff} cells (standard phenotypic definition for T_{reg} and T_{eff} cells throughout) purified from C57BL/6 mice, and in $Foxp3^{+}$ T_{reg} cells versus $Foxp3^{-}$ T_{eff} cells sorted from *Foxp3*^{gfp} knock-in mice¹³ (Fig. 1a). We then investigated whether any of the IL-12 cytokine family α chains (*Il12a*, *Il23a* (which encodes p19), *Il27*) were expressed in peripheral T_{reg} cells, and found that *Il12a* was the only IL-12 family α chain expressed in T_{reg} cells (Fig. 1b). Further analysis confirmed that T_{eff} cells ($CD45RB^{hi}$ or $CD45RB^{lo}$) express negligible amounts of *Ebi3* or *Il12a* mRNA, suggesting restriction to $Foxp3^{+}$ T_{reg} amongst $CD4^{+}$ T cells (Supplementary Fig. 2). Intracellular *Ebi3* expression in resting wild-type T_{reg} cells, but not wild-type T_{eff} or *Ebi3*^{−/−} T_{reg} cells, was

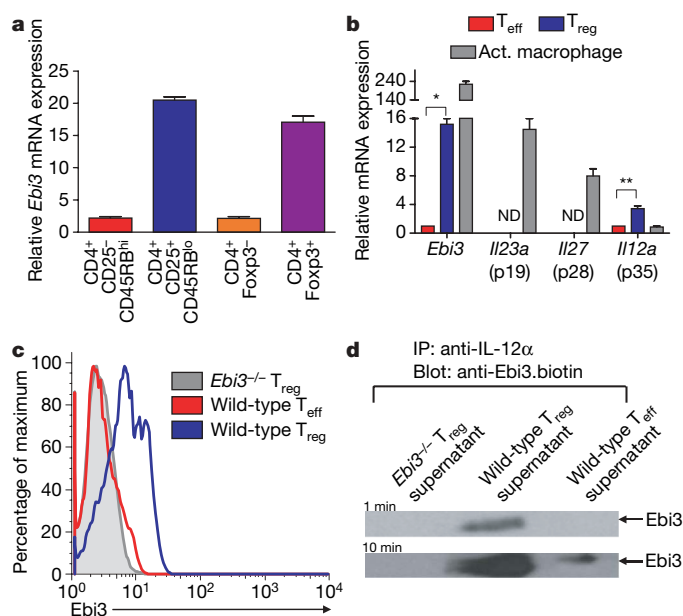


Figure 1 | T_{reg} -restricted expression of *Ebi3* and *Il12a*. RNA was extracted and cDNA generated from FACS-purified T_{eff} or T_{reg} cells. **a**, **b**, Quantitative real-time PCR analysis (asterisk, $P = 0.008$; double asterisk, $P = 0.06$). Act. macrophage, lipopolysaccharide-activated macrophage; ND, not determined. **c**, After brefeldin A treatment, unstimulated T_{eff} and T_{reg} cells were stained with the anti-*Ebi3* monoclonal antibody and were analysed by flow cytometry. **d**, Lysates and supernatants from unstimulated T_{eff} and T_{reg} cells were immunoprecipitated (IP) with an anti-IL-12 α monoclonal antibody, resolved on an SDS-PAGE gel and blotted with anti-*Ebi3* monoclonal antibody. Two exposure times are shown. Data represent the mean \pm s.e.m. of four (**a**), two (**b**), three (**c**) and two (**d**) independent experiments.

¹Department of Immunology, ²Animal Resources Center, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ³Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴eBioscience, San Diego, California 92121, USA.

demonstrated by flow cytometry using three different Ebi3-specific monoclonal antibodies (Fig. 1c and Supplementary Fig. 3). Finally, immunoblot analysis confirmed the co-immunoprecipitation of Ebi3 with IL-12 α in supernatants from resting T_{reg} cells, but not T_{eff} or Ebi3^{-/-} T_{reg} cells (Fig. 1d and Supplementary Fig. 4). These data demonstrate the preferential secretion of a novel Ebi3–IL-12 α heterodimeric cytokine by T_{reg} cells. Ebi3 and IL-12 α had been shown previously to associate in transfected cells; however, no function was ascribed to this heterodimer¹⁴. Analysis of other haematopoietic populations suggested that there was a significant but low level of Ebi3 and IL12a expression in $\gamma\delta$ and CD8⁺ T cells (Supplementary Fig. 2). Given that T_{reg} cells require activation through their T-cell antigen receptor (TCR) to exert their suppressive activity^{15,16}, we assessed how Ebi3 and IL12a mRNA levels were altered after T_{reg} cell activation in the absence or presence of T_{eff} cells. Both Ebi3 and IL12a mRNA were reduced significantly after combined anti-CD3 and anti-CD28 stimulation, but were markedly upregulated (234- and 740-fold, respectively) in T_{reg} cells recovered from an *in vitro* T_{reg} assay, and thus in the process of active suppression (Supplementary Fig. 2). These data suggest that production of the Ebi3–IL-12 α heterodimeric cytokine by T_{reg} cells is potentiated during active suppression of T_{eff} cells.

IL12a mRNA was comparable in Foxp3⁺CD4⁺ and Foxp3⁺CD4⁺ thymocytes, in contrast to the differential expression seen in mature peripheral T cells (Supplementary Fig. 5a; see also Fig. 1b and Supplementary Fig. 2). However, Ebi3 mRNA was present in CD4⁺Foxp3⁺ thymocytes, but was essentially absent in CD4⁺CD8⁺ and CD4⁺Foxp3⁻ thymocytes. Thus, we asked if Ebi3 is a downstream target of Foxp3 by transducing purified T_{eff} cells with retroviral vectors encoding Foxp3 plus green fluorescent protein (GFP) or GFP alone. Foxp3-transduced T_{eff} cells exhibited considerably increased levels of the Ebi3 transcript compared with the GFP-alone controls, whereas Foxp3 induced limited expression of IL12a mRNA (Supplementary Fig. 5b). These data provide a mechanistic basis for the restricted secretion of the Ebi3–IL-12 α heterodimer by T_{reg} cells, with Ebi3 being a downstream target of Foxp3.

Neither Ebi3^{-/-} nor IL12a^{-/-} mice have overt autoimmunity or inflammatory disease^{12,17}, and the percentage of T_{reg} cells in these mice and their Foxp3 expression is comparable to that in wild-type mice (Supplementary Fig. 6). It is possible that the lack of a negative regulatory Ebi3–IL-12 α cytokine may be negated by the lack of the proinflammatory cytokines IL-27 and IL-12 in the Ebi3^{-/-} and IL12a^{-/-} mice, respectively. Indeed, Ebi3^{-/-} mice are more susceptible to leishmaniasis¹⁸. Likewise, IL12a^{-/-}, distinct from IL12b^{-/-} mice, are more susceptible to *Helicobacter*-induced colitis¹⁹, *Leishmania major* infection¹⁷, experimental autoimmune encephalomyelitis²⁰ and collagen-induced arthritis²¹. This prompted us to assess the functional

importance of Ebi3–IL-12 α expression using purified T_{reg} cells from Ebi3^{-/-} and IL12a^{-/-} mice^{12,17}. We first assessed the ability of wild-type, Ebi3^{-/-} and IL12a^{-/-} T_{reg} cells to suppress the proliferation of wild-type T_{eff} cells *in vitro*. Whereas wild-type T_{reg} cells exhibited potent titratable regulatory activity, the suppressive capacity of Ebi3^{-/-} and IL12a^{-/-} T_{reg} cells was significantly reduced, regardless of the level of T_{eff} proliferation (Fig. 2a and Supplementary Fig. 7).

T_{reg} cells have been shown to control the homeostatic expansion of T_{eff} cells in a lymphopenic, recombination activating gene 1 (*Rag1*)^{-/-} environment^{5,22}. Purified wild-type Thy 1.1⁺ T_{eff} cells, either alone or in the presence of wild-type, Ebi3^{-/-} or IL12a^{-/-} Thy 1.2⁺ T_{reg} cells, were adoptively transferred into *Rag1*^{-/-} mice, and splenic T-cell numbers were determined seven to ten days later. In the presence of wild-type T_{reg} cells, T_{eff} cell expansion was significantly reduced (Fig. 2b and Supplementary Fig. 8). In contrast, minimal reduction in wild-type T_{eff} cell expansion was observed in the presence of either Ebi3^{-/-} or IL12a^{-/-} T_{reg} cells. In both *in vitro* and *in vivo* assays, wild-type and Ebi3^{-/-} T_{reg} cell expansion was comparable (Supplementary Figs 7d and 8b, d).

T_{reg} cells have also been shown to control colitis in mice (resembling inflammatory bowel disease, IBD) that is initiated experimentally by transferring naive T cells into *Rag1*^{-/-} recipients²³. Severity of disease is monitored clinically (by weight loss) and histologically (using a semiquantitative grading scheme to score pathology). Recovery from disease, marked by weight gain and decreased histopathology, is observed only in mice that received purified T_{reg} cells approximately four weeks after the initial T_{eff} cell transfer²⁴. We used this recovery model of IBD to test the functionality of Ebi3^{-/-} and IL12a^{-/-} T_{reg} cells *in vivo*. After wild-type T_{eff}-recipient *Rag1*^{-/-} mice developed clinical symptoms of IBD (approximately four weeks), they received wild-type, Ebi3^{-/-} or IL12a^{-/-} T_{reg} cells and were monitored daily. Mice that received wild-type T_{reg} cells were noticeably

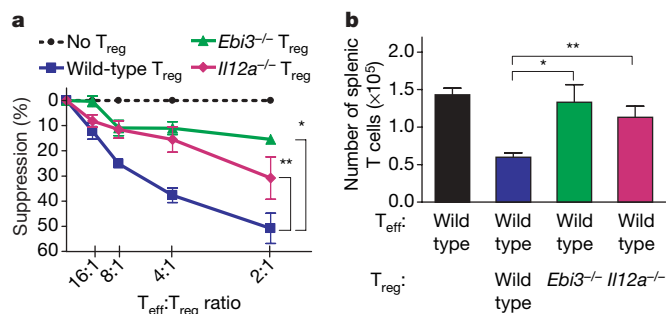


Figure 2 | Ebi3^{-/-} and IL12a^{-/-} T_{reg} cells are functionally defective. **a**, FACS-purified T_{reg} cells were mixed at different ratios with antigen-presenting cells, naive wild-type T_{eff} cells and anti-CD3 for 72 h. Proliferation was determined by [³H]-thymidine incorporation (asterisk, $P = 0.0002$; double asterisk, $P = 0.008$). **b**, T_{eff} cells alone or with wild-type, Ebi3^{-/-} or IL12a^{-/-} T_{reg} cells were injected into *Rag1*^{-/-} mice. Seven days after transfer, mice were euthanized and splenic T-cell numbers determined by flow cytometry (asterisk, $P = 0.002$; double asterisk, $P = 0.02$). Data represent the mean \pm s.e.m. of five (**a**) or three (**b**) independent experiments.

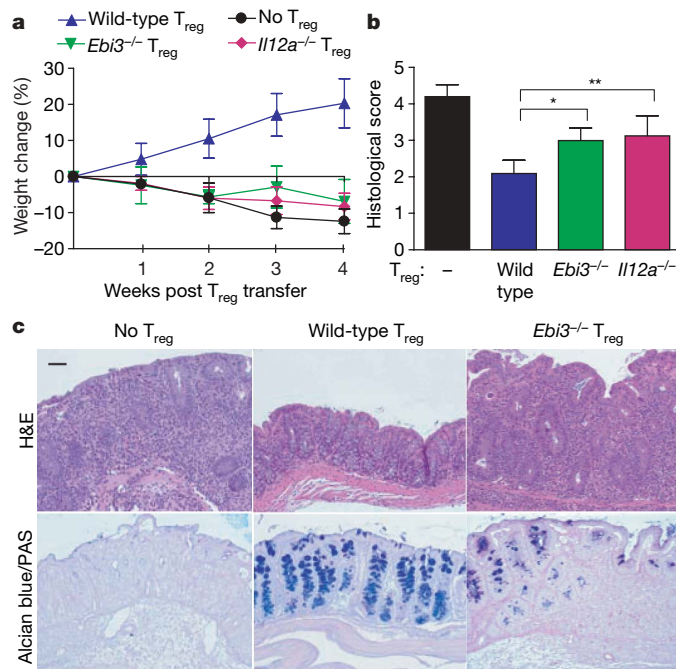


Figure 3 | Ebi3^{-/-} and IL12a^{-/-} T_{reg} cells fail to cure IBD. *Rag1*^{-/-} mice received T_{eff} cells via the tail vein. After 3–4 weeks, mice developed clinical symptoms of IBD and were given wild-type, Ebi3^{-/-} or IL12a^{-/-} T_{reg} cells. **a**, Percentage weight change after T_{reg} transfer. **b**, Colonic histology scores of experimental mice are shown. Data represent mean \pm s.e.m. of 8–11 mice per group from 4 independent experiments (asterisk, $P = 0.02$; double asterisk, $P = 0.05$). **c**, Histological sections of representative colons were stained with haematoxylin and eosin (H&E) or Alcian blue/PAS (Periodic acid Schiff) as indicated (see Supplementary Fig. 9 for full analysis). Scale bar represents 50 μ m.

Inflammatory bowel disease. The recovery model of IBD was used as described previously²⁴. In summary, purified T_{eff} cells were transferred into *Rag1*^{-/-} mice to induce IBD. On clinical signs of disease, mice were given T_{reg} cells. Spleen and lymph node cellularity and colonic pathology were determined eight weeks after the initial T-cell transfer.

Transfections and transductions. HEK293T cells were transfected with an IL-35 plasmid using TransIT transfection reagent. Activated T_{eff} cells were retrovirally transduced by spin transduction. Transduced cells were used in *in vitro* T_{reg} assays.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 31 July; accepted 26 September 2007.

- Sakaguchi, S. *et al.* Immunologic tolerance maintained by CD25⁺ CD4⁺ regulatory T cells: their common role in controlling autoimmunity, tumor immunity, and transplantation tolerance. *Immunol. Rev.* **182**, 18–32 (2001).
- Shevach, E. M. *et al.* The lifestyle of naturally occurring CD4⁺ CD25⁺ Foxp3⁺ regulatory T cells. *Immunol. Rev.* **212**, 60–73 (2006).
- Xystrakis, E., Boswell, S. E. & Hawrylowicz, C. M. T regulatory cells and the control of allergic disease. *Expert Opin. Biol. Ther.* **6**, 121–133 (2006).
- Coombes, J. L., Robinson, N. J., Maloy, K. J., Uhlig, H. H. & Powrie, F. Regulatory T cells and intestinal homeostasis. *Immunol. Rev.* **204**, 184–194 (2005).
- Annacker, O., Pimenta-Araujo, R., Buren-Defranoux, O. & Bandeira, A. On the ontogeny and physiology of regulatory T cells. *Immunol. Rev.* **182**, 5–17 (2001).
- Belkaid, Y., Blank, R. B. & Suffia, I. Natural regulatory T cells and parasites: a common quest for host homeostasis. *Immunol. Rev.* **212**, 287–300 (2006).
- Rouse, B. T., Sarangi, P. P. & Suvas, S. Regulatory T cells in virus infections. *Immunol. Rev.* **212**, 272–286 (2006).
- Kretschmer, K., Apostolou, I., Jaeckel, E., Khazaie, K. & von Boehmer, H. Making regulatory T cells with defined antigen specificity: role in autoimmunity and cancer. *Immunol. Rev.* **212**, 163–169 (2006).
- Gavin, M. A. *et al.* Foxp3-dependent programme of regulatory T-cell differentiation. *Nature* **445**, 771–775 (2007).
- Devergne, O. *et al.* A novel interleukin-12 p40-related protein induced by latent Epstein-Barr virus infection in B lymphocytes. *J. Virol.* **70**, 1143–1153 (1996).
- Pflanz, S. *et al.* IL-27, a heterodimeric cytokine composed of EBI3 and p28 protein, induces proliferation of naive CD4⁺ T cells. *Immunity* **16**, 779–790 (2002).
- Nieuwenhuis, E. E. *et al.* Disruption of T helper 2-immune responses in Epstein-Barr virus-induced gene 3-deficient mice. *Proc. Natl Acad. Sci. USA* **99**, 16951–16956 (2002).
- Fontenot, J. D. *et al.* Regulatory T cell lineage specification by the forkhead transcription factor Foxp3. *Immunity* **22**, 329–341 (2005).
- Devergne, O., Birkenbach, M. & Kieff, E. Epstein-Barr virus-induced gene 3 and the p35 subunit of interleukin 12 form a novel heterodimeric hematopoietin. *Proc. Natl Acad. Sci. USA* **94**, 12041–12046 (1997).
- Thornton, A. M. & Shevach, E. M. CD4⁺CD25⁺ immunoregulatory T cells suppress polyclonal T cell activation *in vitro* by inhibiting interleukin 2 production. *J. Exp. Med.* **188**, 287–296 (1998).
- Takahashi, T. *et al.* Immunologic self-tolerance maintained by CD25⁺ CD4⁺ naturally anergic and suppressive T cells: induction of autoimmune disease by breaking their anergic/suppressive state. *Int. Immunol.* **10**, 1969–1980 (1998).
- Mattner, F. *et al.* Genetically resistant mice lacking interleukin-12 are susceptible to infection with *Leishmania major* and mount a polarized Th2 cell response. *Eur. J. Immunol.* **26**, 1553–1559 (1996).
- Zahn, S. *et al.* Impaired Th1 responses in mice deficient in Epstein-Barr virus-induced gene 3 and challenged with physiological doses of *Leishmania major*. *Eur. J. Immunol.* **35**, 1106–1112 (2005).
- Kullberg, M. C. *et al.* IL-23 plays a key role in *Helicobacter hepaticus*-induced T cell-dependent colitis. *J. Exp. Med.* **203**, 2485–2494 (2006).
- Gran, B. *et al.* IL-12p35-deficient mice are susceptible to experimental autoimmune encephalomyelitis: evidence for redundancy in the IL-12 system in the induction of central nervous system autoimmune demyelination. *J. Immunol.* **169**, 7104–7110 (2002).
- Murphy, C. A. *et al.* Divergent pro- and antiinflammatory roles for IL-23 and IL-12 in joint autoimmune inflammation. *J. Exp. Med.* **198**, 1951–1957 (2003).
- Workman, C. J. & Vignali, D. A. A. Negative regulation of T cell homeostasis by LAG-3 (CD223). *J. Immunol.* **174**, 688–695 (2004).
- Izcue, A., Coombes, J. L. & Powrie, F. Regulatory T cells suppress systemic and mucosal immune activation to control intestinal inflammation. *Immunol. Rev.* **212**, 256–271 (2006).
- Mottet, C., Uhlig, H. H. & Powrie, F. Cutting edge: cure of colitis by CD4⁺CD25⁺ regulatory T cells. *J. Immunol.* **170**, 3939–3943 (2003).
- Brombacher, F., Kastelein, R. A. & Alber, G. Novel IL-12 family members shed light on the orchestration of Th1 responses. *Trends Immunol.* **24**, 207–212 (2003).
- Fuss, I. J. *et al.* Both IL-12p70 and IL-23 are synthesized during active Crohn's disease and are down-regulated by treatment with anti-IL-12 p40 monoclonal antibody. *Inflamm. Bowel Dis.* **12**, 9–15 (2006).
- Schrader, J. W. Interleukin is as interleukin does. *Trends Immunol.* **23**, 573–574 (2002).
- Hori, S., Nomura, T. & Sakaguchi, S. Control of regulatory T cell development by the transcription factor Foxp3. *Science* **299**, 1057–1061 (2003).
- Huang, C. T. *et al.* Role of LAG-3 in regulatory T cells. *Immunity* **21**, 503–513 (2004).
- Kaplan, D. Autocrine secretion and the physiological concentration of cytokines. *Immunol. Today* **17**, 303–304 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A.-M. Clark for molecular analysis; K. Forbes for colony management; J. Rogers for FACS; D. Finkelstein and T. Xu for Affymetrix GeneChip data analysis; staff at the St Jude Hartwell Center for Affymetrix GeneChip probing, oligo synthesis and DNA sequencing; staff at the St Jude ARC Histology Laboratory and Animal Husbandry Unit, and the Flow Cytometry and Cell Sorting Shared Resource facility for MACS; J. Fisher for immunoprecipitation and western blot advice; and P. Just, G. Li and D. Mitchell for Ebi3 intracellular staining and western blotting reagents. We also thank D. Green, D. Pardoll and T. Geiger for their review of the manuscript; members of the Vignali laboratory for discussions; H. von Boehmer and D. Pardoll for the 6.5 TCR transgenic mice; and A. Rudensky for the *Foxp3*^{gfp} knock-in mice. This work was supported by the National Institutes of Health (D.A.A.V., R.S.B. and T.T.K.), by a Cancer Center Support CORE grant and the American Lebanese Syrian Associated Charities (ALSAC) (D.A.A.V.), by a St Jude Gephhardt Postdoctoral Fellowship and an Individual NRSA (L.W.C.), and by the American Liver Foundation (T.T.K.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to D.A.A.V. (dario.vignali@stjude.org).

METHODS

Mice. *Ebi3*^{-/-} mice¹² (C57BL/6; F₀, now 98.83% C57BL/6 by microsatellite analysis, Charles River) were initially provided by R. Blumberg and T. Kuo; they were subsequently obtained from our own breeding colony, which was re-derived at Charles River Breeding Laboratories and housed at St Jude Children's Research Hospital. *Foxp3*^{gfp} mice¹³ (C57BL/6; F₇, now 95.32% C57BL/6 by microsatellite analysis, Charles River) were provided by A. Rudensky. Haemagglutinin-specific (¹¹⁰SPERFEIFPKE¹²⁰) H-2E^d-restricted 6.5 TCR-transgenic mice³¹ (B10.D2, backcrossed for ten generations) were provided by H. von Boehmer and D. Pardoll. *Il12a*^{-/-} (ref. 17) and C57BL/6 mice were purchased from the Jackson Laboratory. All animal experiments were performed in American Association for the Accreditation of Laboratory Animal Care-accredited, specific-pathogen-free, *Helicobacter*-free facilities in the St Jude Animal Resource Center following national, state and institutional guidelines. Animal protocols were approved by the St Jude Animal Care and Use Committee.

Flow cytometric analysis, intracellular staining and cell sorting. T_{eff} (CD4⁺CD25⁻CD45RB^{hi}) and T_{reg} (CD4⁺CD25⁺CD45RB^{lo}) cells from the spleens and lymph nodes of C57BL/6 or knockout age-matched mice were positively sorted by FACS. After red blood cell lysis with Gey's solution, cells were stained with antibodies against CD4, CD25 and CD45RB (eBioscience) and sorted on a MoFlo (Dako) or Reflection (i-Cyt). Thymocytes from *Foxp3*^{gfp} mice were positively sorted by FACS using antibodies against CD4 and CD8 (Becton Dickinson). CD4⁺ T_{eff} cells from 6.5 TCR transgenic mice were negatively purified by MACS separation (AutoMACS, Miltenyi Biotec). Splenocytes were stained with biotinylated anti-B220, anti-Gr1 (also known as Ly6g), anti-Mac1 (Itgam), anti-TER119 (Ly76), anti-pan NK (Itga2), anti-CD25 and anti-CD8, followed by streptavidin-coupled magnetic beads. Flow cytometric analysis was performed as detailed previously³², using a FACSCalibur (Becton Dickinson). Intracellular staining for Ebi3 was performed with three monoclonal antibodies (4D12, 4H1 and 30A1, provided by D. Sehy). T_{eff} and T_{reg} cells from *Ebi3*^{-/-} or C57BL/6 mice were isolated by FACS as described previously. After 28 h in culture in the absence of stimuli, the cells were incubated with 1:100 Golgi plug containing brefeldin A (BD Bioscience) for 8 h. The cells were fixed and permeabilized with the cytofix/cytoperm kit (BD Bioscience), stained with Alexafluor 647-conjugated, rat anti-mouse Ebi3 monoclonal antibodies (eBioscience) and analysed by flow cytometry.

RNA, complementary DNA and quantitative real-time PCR. Purified T cells from *Foxp3*^{gfp} mice were activated with anti-CD3- and anti-CD28-coated microbeads. Where indicated, T_{reg} cells were activated in the presence of T_{eff} cells at a ratio of 4:1 (T_{eff}:T_{reg}). After 48 h, T_{eff} and T_{reg} cells were re-sorted on the basis of Foxp3 expression for analysis. T-cell RNA was isolated from whole cells using the Qiagen microRNA extraction kit following the manufacturer's instructions. RNA was quantified spectrophotometrically, and complementary DNA was reverse-transcribed using the cDNA archival kit (Applied Biosystems) following the manufacturer's guidelines. TaqMan primers and probes were designed with PrimerExpress software and were synthesized in the St Jude Hartwell Center for Biotechnology and Bioinformatics. *Ebi3* 5' primer, AGCAGCAGCCTCCTAGCCT; *Ebi3* 3' primer, ACGCCTCCGGAGGGTC; *Ebi3* probe, FAM-TGTGGCTGAGCGAATCATCAAGCC. *Il12a* 5' primer, TGGCTACTAGAGAGACTTCTCCACAA; *Il12a* 3' primer, GCACAGGGTCATCATCAAGAG; *Il12a* probe, FAM-AAGAGGGAGCTGCCTGCCCC. *Il27* 5' primer, GGCCAGGYGACAGGAGCC; *Il27* 3' primer, CAGCTTGTA-CCAGAAGCAAGGG; *Il27* probe, FAM-TGGCGGCTCAGCCTGTTGCTG. *Il23a* 5' primer, CCAATGTTTCCCTGACTTTCCA; *Il23a* 3' primer, AAGTGTGGTAGCGAGGAAGCA; *Il23a* probe, FAM-TGGCATCACCTCTCTGA.

The cDNA samples were subjected to 40 cycles of amplification in an ABI Prism 7900 Sequence Detection System instrument according to the manufacturer's protocol. Quantification of relative mRNA expression was determined by the comparative CT (critical threshold) method as described in the ABI User Bulletin number 2 (<http://docs.appliedbiosystems.com/pebi/docs/04303859.pdf>), whereby the amount of target mRNA, normalized to endogenous β-actin expression, is determined by the formula 2^{-ΔΔCT}.

Immunoprecipitation and western blotting. Immunoprecipitation and immunoblotting were performed as described previously^{33,34}. Sorted T_{eff} and T_{reg} cells (3 × 10⁶ cells per lane) were cultured for 36 h in the absence of stimuli. In summary, cells were lysed with lysis buffer containing 0.1% Tween 20, 50 mM HEPES, 150 mM NaCl, 1 mM EDTA, 2.5 mM EGTA and one complete protease inhibitor tablet (Roche) per 50 ml lysis buffer, pH 7.4, on ice for 30 min. This was followed by centrifugation at 15,000g for 10 min. Cleared lysates or cellular supernatants were immunoprecipitated with anti-mouse IL-12α (p35) antibody (clone C18.2, eBioscience), pre-coupled to protein G-Sepharose beads. Immunoprecipitates were resolved by SDS-PAGE (Invitrogen Life

Technologies), and blots were probed with a biotinylated monoclonal anti-mouse Ebi3 antibody (30A1, eBioscience). Blots were developed using chemiluminescence (Amersham Biosciences) and autoradiography.

In vitro suppression assays for natural T_{reg} cells. T_{reg} cell assays were performed as described previously with some modifications²⁹. T_{eff} and T_{reg} cells from wild-type C57BL/6, *Ebi3*^{-/-} and *Il12a*^{-/-} mice were sorted by FACS. *In vitro* T_{reg} function was measured by culturing 2.5 × 10⁴ T_{eff} cells with 1 × 10⁵ irradiated (3,000 rads) C57BL/6 splenocytes as antigen-presenting cells, 1 μg ml⁻¹ anti-CD3, and titrations of wild-type, *Ebi3*^{-/-} and *Il12a*^{-/-} T_{reg} cells. Cells were pre-incubated for 10 min in 4 μM CFSE in PBS plus 0.1% BSA before culture for CFSE dilution analysis. Where indicated, irradiated transfected 293T cells (5,000 rads) or culture medium from transfected 293T cells (dialysed against media and filtered) were added to cultures in the place of T_{reg} cells. In these assays, T_{eff} cells were activated with anti-CD3- and anti-CD28-coated microbeads. Cultures were pulsed with 1 μCi [³H]-thymidine for the final 8 h of the 72 h assay, and were harvested with a Packard Micromate cell harvester. Counts per minute were determined using a Packard Matrix 96 direct counter (Packard Biosciences).

Adoptive transfer for homeostatic expansion. Homeostasis assays were performed as described previously²². In summary, naive T_{eff} (CD4⁺CD25⁻CD45RB^{hi}) and T_{reg} (CD4⁺CD25⁺CD45RB^{lo}) cells were isolated by FACS from pooled spleens and lymph nodes of C57BL/6, B6.PL, *Ebi3*^{-/-} or *Il12a*^{-/-} mice. T_{eff} cells (2 × 10⁶) with or without T_{reg} cells (5 × 10⁵) were stained, at room temperature, with 4 μM CFSE in PBS plus 0.1% BSA for 10 min. After two washes with PBS, cells were resuspended in 0.5 ml of PBS plus 2% FBS, and were injected intravenously through the tail vein into *Rag1*^{-/-} mice. Mice were euthanized seven days post transfer, and splenocytes were counted, stained and analysed by flow cytometry using antibodies against Thy1.1, Thy1.2 (BD Bioscience), CD4, CD25 and Foxp3 (eBioscience). Where indicated, congenic mice were used to determine numbers of Thy 1.1⁺ T_{eff} cells and Thy 1.2⁺ T_{reg} cells. For each group, 8–12 mice were analysed.

Inflammatory bowel disease model. The recovery model of IBD was used, with some modifications³⁵. In summary, *Rag1*^{-/-} mice were injected intravenously with 4 × 10⁵ wild-type CD4⁺CD25⁻CD45RB^{hi} T_{eff} cells to induce IBD. On clinical signs of disease, approximately four weeks post transfer, mice were divided into three groups. Some mice were euthanized to determine the severity of disease. Experimental groups received 7.5 × 10⁵ T_{reg} cells from wild-type, *Ebi3*^{-/-} or *Il12a*^{-/-} mice by means of intraperitoneal injection. Where indicated, T_{reg} cells alone were also injected, independent of T_{eff} cells. All mice were weighed weekly and were euthanized 32 days post transfer (~eight weeks after the initial T_{eff} transfer). Colons were sectioned, fixed in 10% neutral buffered formalin and processed routinely, and 4-μm sections cut and stained with H&E or Alcian blue/Periodic acid Schiff. CD3⁺ T cells were visualized using a goat anti-CD3 polyclonal antisera (Santa Cruz) and diaminobenzidine chromagen with haematoxylin as a counterstain. Foxp3⁺ cells were visualized with rat anti-Foxp3 clone FJK-16S antibody (eBioscience). Pathology of the large intestine was scored blindly using a semiquantitative scale of zero to five as described previously³⁶. In summary, grade 0 was assigned when no changes were observed; grade 1, minimal inflammatory infiltrates present in the lamina propria with or without mild mucosal hyperplasia; grade 2, mild inflammation in the lamina propria with occasional extension into the submucosa, focal erosions, minimal to mild mucosal hyperplasia and minimal to moderate mucin depletion; grade 3, mild to moderate inflammation in the lamina propria and submucosa occasionally transmural with ulceration and moderate mucosal hyperplasia and mucin depletion; grade 4 marked inflammatory infiltrates commonly transmural with ulceration, marked mucosal hyperplasia and mucin depletion, and multifocal crypt necrosis; grade 5, marked transmural inflammation with ulceration, widespread crypt necrosis and loss of intestinal glands. Spleen and mesenteric lymph nodes were also removed, and cells were counted and subsequently stained for CD4 and Foxp3 to determine T_{eff} (CD4⁺Foxp3⁻) and T_{reg} (CD4⁺Foxp3⁺) numbers in spleen and mesenteric lymph nodes by flow cytometry.

Transfection of HEK293T cells and retroviral transduction of T_{eff} cells. IL-35 constructs were generated by recombinant PCR as described^{32,37}, and were cloned into either pMIG, a mouse stem-cell-virus-based retroviral vector containing an internal ribosomal entry site (IRES) and coding for GFP, or pPIGneo, a pCNeo-based vector (Promega) used for protein expression in HEK293T cells that we have modified to include an IRES–GFP cassette. In the 'native' IL-35 construct, *Ebi3* and *Il12a* are linked in a single open reading frame by means of a 'self-cleaving' 2A peptide (*Ebi3*-2A-*Il12a*) that generates stoichiometric levels of two separate proteins that then pair in the endoplasmic reticulum before secretion^{37,38–40}. In the 'single-chain' IL-35 construct, Ebi3 and IL-12α are expressed as a single-chain protein connected by means of a flexible (GGGS)₄ linker (*Ebi3*-(GGGS)₄-IL-12α), an approach that has been used successfully to generate biologically active recombinant IL-27 (ref. 41). Constructs containing *Ebi3* and *Il12a* alone were used as controls. HEK293T cells were transfected using 10 μg

plasmid per 5×10^6 cells using TransIT transfection reagent (Mirus). Cells were sorted for equivalent GFP expression and were cultured for 36 h to facilitate protein secretion. Dialysed, filtered supernatant from cells was mixed at indicated ratios with anti-CD3/CD28-coated sulphate latex beads and T_{eff} cells in a proliferation assay. For transduction of T_{eff} cells, retrovirus was produced as described^{42–44} and was then filtered with a $0.45 \mu\text{m}$ filter, layered onto activated T_{eff} cells (at 2.5×10^6 cells per ml) in the presence polybrene ($6 \mu\text{g ml}^{-1}$), and spun at 2,000g for 90 min on two sequential days. For ectopic expression of constructs in $6.5 T_{eff}$ cells, cells were allowed to rest for ten days in 2 ng ml^{-1} IL-2 and sorted to obtain cells with the highest 30% GFP expression. Sorted cells were titrated into an *in vitro* T_{reg} assay with antigen-presenting cells, $10 \mu\text{g ml}^{-1}$ haemagglutinin 110–120 peptide and naive $6.5 CD4^+CD25^- T_{eff}$ cells.

31. Jooss, K., Gjatta, B., Danos, O., von Boehmer, H. & Sarukhan, A. Regulatory function of *in vivo* anergized $CD4^+$ T cells. *Proc. Natl Acad. Sci. USA* **98**, 8738–8743 (2001).
32. Vignali, D. A. & Vignali, K. M. Profound enhancement of T cell activation mediated by the interaction between the T cell receptor and the D3 domain of CD4. *J. Immunol.* **162**, 1431–1439 (1999).
33. Li, N., Workman, C. J., Martin, S. M. & Vignali, D. A. A. Biochemical analysis of the regulatory T cell protein lymphocyte activation gene-3 (LAG-3; CD223). *J. Immunol.* **173**, 6806–6812 (2004).
34. Li, N. *et al.* Metalloproteases regulate T cell proliferation and effector function via LAG-3. *EMBO J* **26**, 494–504 (2007).
35. Mottet, C., Uhlig, H. H. & Powrie, F. Cutting edge: cure of colitis by $CD4^+CD25^+$ regulatory T cells. *J. Immunol.* **170**, 3939–3943 (2003).
36. Asseman, C., Mauze, S., Leach, M. W., Coffman, R. L. & Powrie, F. An essential role for interleukin 10 in the function of regulatory T cells that inhibit intestinal inflammation. *J. Exp. Med.* **190**, 995–1004 (1999).
37. Szymczak-Workman, A. L., Vignali, K. M. & Vignali, D. A. A. in *Gene Transfer: Delivery and Expression* (eds Friedmann, T. & Rossi, J.) 137–147 (Cold Spring Harbor Laboratory, Cold Spring Harbor, 2006).
38. Szymczak, A. *et al.* Correction of multi-gene deficiency *in vivo* using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nature Biotechnol.* **22**, 589–594 (2004).
39. Szymczak, A. & Vignali, D. A. A. Development of 2A peptide-based strategies in the design of multicistronic vectors. *Exp. Opin. Biol. Ther.* **5**, 627–638 (2005).
40. Holst, J., Vignali, K. M., Burton, A. R. & Vignali, D. A. A. Rapid analysis of T cell selection and function *in vivo* using T cell receptor retrogenic mice. *Nature Methods* **3**, 191–197 (2006).
41. Hisada, M. *et al.* Potent antitumour activity of interleukin-27. *Cancer Res.* **64**, 1152–1156 (2004).
42. Holst, J. *et al.* Generation of T cell receptor retrogenic mice. *Nature Protocols* **1**, 406–417 (2006).
43. Persons, D. A. *et al.* Retroviral-mediated transfer of the green fluorescent protein gene into murine haematopoietic cells facilitates scoring and selection of transduced progenitors *in vitro* and identification of genetically modified cells *in vivo*. *Blood* **90**, 1777–1786 (1997).
44. Persons, D. A., Mehaffey, M. G., Kaleko, M., Nienhuis, A. W. & Vanin, E. F. An improved method for generating retroviral producer clones for vectors lacking a selectable marker gene. *Blood Cells Mol. Dis.* **24**, 167–182 (1998).

CORRIGENDUM

The inhibitory cytokine IL-35 contributes to regulatory T-cell function

Lauren W. Collison, Creg J. Workman, Timothy T. Kuo, Kelli Boyd, Yao Wang, Kate M. Vignali, Richard Cross, David Sehy, Richard S. Blumberg & Dario A. A. Vignali

Nature **450**, 566–569 (22 November 2007)

In the Methods, 'Il23a 5' primer, CCAATGTTTCCCTGACTTTCCA; Il23a 3' primer, AAGTGTGGTAGCGAGGAAGCA; Il23a probe, FAM-TGGCATCACCTCTCTGA' should read 'Il23a 5' primer, ATCCAGTGTGAAGATGGTTGTGA; Il23a 3' primer, GCAAGCAG-AACTGGCTGTTG; Il23a probe, FAM-CCACAAGGACTCAAGG'.

LETTERS

A SNARE–adaptor interaction is a new mode of cargo recognition in clathrin-coated vesicles

Sharon E. Miller¹, Brett M. Collins², Airlie J. McCoy¹, Margaret S. Robinson¹ & David J. Owen¹

Soluble NSF attachment protein receptors (SNAREs) are type II transmembrane proteins that have critical roles in providing the specificity and energy for transport-vesicle fusion and must therefore be correctly partitioned between vesicle and organelle membranes^{1–3}. Like all other cargo, SNAREs need to be sorted into the forming vesicles by direct interaction with components of the vesicles' coats. Here we characterize the molecular details governing the sorting of a SNARE into clathrin-coated vesicles, namely the direct recognition of the three-helical bundle H_{abc} domain of the mouse SNARE Vti1b by the human clathrin adaptor epsinR (EPNR, also known as CLINT1). Structures of each domain and of their complex show that this interaction (dissociation constant 22 μ M) is mediated by surface patches composed of approximately 15 residues each, the topographies of which are dependent on each domain's overall fold. Disruption of the interface with point mutations abolishes the interaction *in vitro* and causes Vti1b to become relocalized to late endosomes and lysosomes. This new class of highly specific, surface–surface interaction between the clathrin coat component and the cargo is distinct from the widely observed binding of short, linear cargo motifs by the assembly polypeptide (AP) complex and GGA adaptors⁴ and is therefore not vulnerable to competition from standard motif-containing cargoes for incorporation into clathrin-coated vesicles. We propose that conceptually similar but mechanistically different interactions will direct the post-Golgi trafficking of many SNAREs.

The cytoplasmic domains of SNAREs all contain a 16-turn 'SNARE' helix, amino-terminal to which is a region that varies from a 20-residue unstructured peptide to a 100-residue helical H_{abc} domain or a 150-residue longin domain (reviewed in ref. 1). The 36 mammalian SNAREs can be defined as either Q- or R-SNAREs, depending on the residue at layer 0 in the SNARE helix². Specific combinations of three Q-SNAREs from one membrane and one R-SNARE from another interact to form a trans-SNAREpin complex that brings the membranes into close apposition, and therefore facilitates their fusion³. To ensure that a vesicle can fuse with its target membrane, sufficient amounts of the correct SNAREs must be incorporated into the vesicle during its formation, and cognate SNAREs must also be already present on the target membrane. For vesicle trafficking to continue in the cell and for membrane identity to be maintained, some SNAREs must be recycled back to their steady-state locations whereas others are retained in their current organelle; in addition, any that have become mislocalized from their normal trafficking routes need to be retrieved.

Post-Golgi vesicular transport is mediated mainly by clathrin-coated vesicles (CCVs), the coats of which are composed of an outer clathrin scaffold linked to the membrane by clathrin adaptors. Cargo selection is achieved by the binding of short, linear motifs in their cytoplasmic portions by a subset of clathrin adaptors. These general

cargo/clathrin adaptors include AP (assembly polypeptide) complexes that recognize Yxx Φ (where Φ is a bulky hydrophobic residue) and (D/E)xxxLL, GGAs that recognize DxxLL, and family-specific cargo/clathrin adaptors such as mammalian Ldlrap1 (also known as Arh) and Dab2 that recognize FxNPxY motifs on low-density lipoprotein family members (for a review of clathrin adaptors, see ref. 4). Ubiquitin, a small transplantable structural motif that can be coupled to any cargo by means of a suitable lysine residue, can also be recognized with low affinity by GGAs and probably by epsins 1 and 2 (reviewed in ref. 5). However, most SNAREs lack the known adaptor-binding trafficking motifs (although a few, including vesicle-associated membrane protein 4, VAMP4 (ref. 6), do possess potential di-leucine motifs that can bind AP complexes and, therefore, could identify them for retrieval⁷), and it is hard to envisage how frequently used motifs would provide the necessary highly specific targeting of a very few SNAREs to a certain vesicle to ensure that this vesicle will fuse only with its correct target compartment. Therefore, the question arises as to how different SNAREs are selected for incorporation into different CCVs. One possible solution would be for SNAREs to interact directly with components of a CCV that are not classified as conventional cargo adaptors. Several examples have been proposed: these include interactions between the trans-Golgi network (TGN)/endosomal clathrin adaptor EPNR and the Q-SNARE Vti1b^{8,9}, between the endocytic clathrin adaptor AP180 homologue UNC-11 in *C. elegans* and synaptobrevin¹⁰, and between the homologue Yap1801 in *S. cerevisiae* and the Q-SNAREs Sso1 and Sso2 (ref. 11). A second solution would be for the SNARE to interact in a non-standard manner with an already characterized adaptor, as has been proposed for VAMP7 binding to AP3 (ref. 12).

Vti1b forms SNAREpin complexes with the Q-SNAREs syntaxin 7 and syntaxin 8 and either of the R-SNAREs VAMP7 or VAMP8 to drive fusion with endosomal membranes¹³, and forms complexes with syntaxin 6 and syntaxin 7 and an unknown R-SNARE to drive fusion at the TGN¹⁴. The association between EPNR and Vti1b seems to be important for the recruitment of Vti1b into CCVs⁸ and an interaction between EPNR and Vti1b homologues is conserved in yeast⁹ and plants¹⁵. We set out to elucidate the molecular mechanism of CCV recruitment of Vti1b by EPNR to gain insight into to how SNAREs may be recruited into CCVs in general. Previous work has indicated that the epsin N-terminal homology (ENTH) domain of EPNR could bind specifically to a region of Vti1b (residues 1–116) encompassing a putative H_{abc} domain and an unstructured linker sequence before the SNARE helix⁹. We found that a shorter construct (residues 1–96) containing only the putative H_{abc} fold could bind with equal affinity to residues 1–116 (Supplementary Fig. 2), indicating that the interaction was fundamentally different in nature from the interaction between the yeast SNARE Sed5 and the COPII coat, which is mediated by a short linear peptide from the linker

¹University of Cambridge, CIMR, Wellcome Trust/MRC Building, Hills Road, Cambridge, CB2 0XY, UK. ²IMB, University of Queensland, Queensland Bioscience Precinct, Building 80, Services Road, St Lucia, Queensland 4075, Australia.

between the H_{abc} domain and the SNARE helix of Sed5 binding to Sec24 in yeast¹⁶.

The structures of mouse Vti1b(1–96) and the human EPNR ENTH $\Delta\alpha 0$ domain (residues 20–166), in which the putative membrane-embedding amphipathic helix $\alpha 0$ (ref. 17) was deleted to enhance solubility, were each determined by X-ray crystallography at 2 Å resolution (Fig. 1 and Supplementary Figs 3 and 4). These structures were used to identify exposed hydrophobic residues that may mediate interaction between the domains. Mutant constructs of glutathione S-transferase (GST)-tagged Vti1b(1–96) and His₆-Myc-tagged EPNR ENTH were then used in GST pull-down experiments (Fig. 1 and Supplementary Fig. 4). The native fold of all mutants was confirmed by circular dichroism spectroscopy. Initially, one mutation was identified that abolished binding to EPNR—Vti1b(F73S) on Vti1b helix H_c . This identification was supported by the observation that Vti1a, in which this residue is replaced by a serine, and the yeast homologue of Vti1b, Vti1, in which it is replaced by a glutamate (see Fig. 2), are both unable to bind EPNR (Supplementary Fig. 2). Mutation of the spatially close residues Vti1b(S8W) and Vti1b(E12A) on H_a , and Vti1b(E65W) on H_b (Fig. 1), also abrogated the interaction. Of the many hydrophobic surface residues initially mutated in EPNR ENTH, only one, EPNR(F52D), abolished Vti1b binding. Mutating residues around EPNR Phe 52 showed that alterations in EPNR(M53D/Y54D) and EPNR(R96S) also strongly inhibited binding to Vti1b, as did deletion of helix $\alpha 8$ (Fig. 1). Isothermal titration calorimetry (ITC) showed that the interaction between Vti1b and EPNR was exothermic and exhibited a binding dissociation constant

(K_d) of $22 \pm 3 \mu\text{M}$ ($n = 6$). The affinities of all strongly inhibiting point mutants were unmeasurable ($K_d > 300 \mu\text{M}$), whereas removal of helix $\alpha 8$ caused an approximately tenfold decrease in affinity ($K_d = 230 \mu\text{M}$; Supplementary Fig. 4).

The regions of the proteins mapped as being important for the interaction suggested an association between patches on the surfaces of each folded domain, rather than a linear peptide–folded domain interaction as used in other cargo–clathrin–adaptor interactions⁴. To test this hypothesis, we determined the structure of the EPNR ENTH $\Delta\alpha 0$ –Vti1b H_{abc} complex by X-ray crystallography at 2.2 Å resolution (Fig. 2). Analysis of the contents of the crystal lattice using PISA (Protein Interfaces, Surfaces and Assemblies)¹⁸ indicated that the only likely bona fide protein–protein interface was indeed the one we had proposed. Structures of other SNARE H_{abc} domains bound to helical fragments of other proteins^{19,20} show that the bound helix assumes a parallel orientation against H_a and H_b or H_b and H_c , forming a four-helical bundle (Supplementary Fig. 6a). EPNR forms a distinctly different type of interaction with Vti1b H_{abc} , because it binds with its helices oriented almost at right-angles to those of the H_{abc} and uses the surface formed by helices H_a and H_c . The structure of the Vti1b–EPNR complex allows us to propose a model for the assembly of these proteins with respect to the membrane (Fig. 2a). If we assume that the N-terminal residues of EPNR (not included in our construct) fold into an amphipathic helix $\alpha 0$ that embeds in phospholipid bilayers as in epsin 1 (Epn1, ref. 17), then the ENTH domain will lie with its helices parallel to the membrane. Thus, the unstructured carboxy-terminal tail of EPNR, which contains AP1/GGA appendage and clathrin-binding sequences, is free to point away from the membrane to engage the clathrin and adaptor layer. The Vti1b H_{abc} domain will have to fold back on the SNARE helix such that its helices are perpendicular to the membrane.

The elongated interface between the two proteins shows good shape complementarity (value of 0.71, with 1 being a perfect fit²¹) with 1,400 Å² solvent-accessible surface buried by the interaction (approximately 700 Å² on each protein). It involves around 15 residues from helices H_a and H_c on Vti1b, associated with around 15 residues from helix $\alpha 8$ and from the loops between helices $\alpha 2$ and $\alpha 3$ and helices $\alpha 4$ and $\alpha 5$ on EPNR ENTH $\Delta\alpha 0$. The interface contains hydrophobic, hydrogen-bond and electrostatic elements (Fig. 2), explaining the effects of all the non-binding mutants. The validity of the interface was further confirmed by the design of complementary charge-swap mutations EPNR(R146E) and Vti1b(E23R). On their own, each strongly inhibited binding to wild-type versions of the other domain; however, when both mutant domains were used, binding was restored to wild-type levels, confirming the importance of the charged component of the interaction between the two proteins (Figs 1f and 2a). Comparison of the isolated and complexed structures shows that there has been no significant change in the conformation of the EPNR ENTH $\Delta\alpha 0$ domain, but there has been a significant degree of rearrangement of Vti1b H_{abc} residues Vti1b Met 1 to Vti1b Ser 7 (Supplementary Fig. 6b). These residues are altered from a helical structure in the apo domain to a non-helical structure in the bound complex in such a way that they form a cavity into which the EPNR Tyr 54 side-chain fits.

The total buried surface area and strength of interaction described here (1,400 Å² and 22 μM , respectively) are comparable with those of the moderate-affinity (low micromolar K_d) interactions used by clathrin adaptors to bind their cognate cargo by means of short linear peptide motifs; for example, AP2 $\mu 2$ subunits with Yxx Φ motifs (DYQRLN from rat TGN38 (also known as Tgoln2), $K_d = 1.5 \mu\text{M}$; 1,000 Å, refs 4, 22 and 23), and GGA VHS domains with (phosphoserine)DxxLL signals ($K_d = 2–4 \mu\text{M}$; 1,150 Å, refs 1 and 24). In all of these cases there is no pre-formed tertiary structural (that is, fold-dependent) element on the cargo molecule involved in the recognition by the clathrin adaptor, because the sorting motifs are functional both in the unstructured cytosolic portions of many

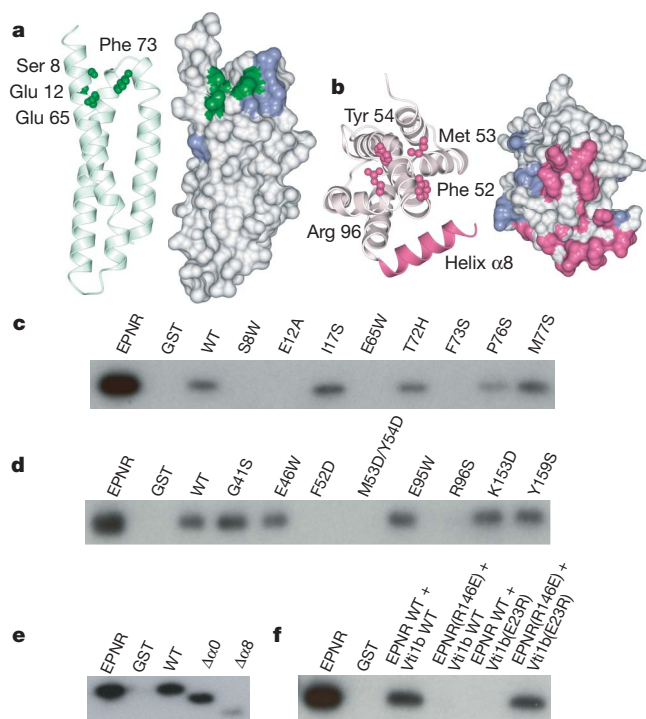


Figure 1 | Mapping the binding sites on the Vti1b H_{abc} domain and the EPNR ENTH $\Delta\alpha 0$ domain on their isolated structures. **a, b**, Ribbon diagram showing the three-helix bundle (H_{abc} domain) of uncomplexed Vti1b (**a**, light green) and of uncomplexed EPNR ENTH $\Delta\alpha 0$ (**b**, pale pink). Surface views are shown in the same orientations. Mutated residues on both representations are coloured green if they affected binding to EPNR, pink if they affected binding to Vti1b H_{abc} and grey-blue if there was no effect (surface views only). **c–f**, Pull-down experiments detecting the binding of EPNR ENTH–Myc constructs to GST–Vti1b by western blotting for the Myc tag. **c**, The effect of mutations in GST–Vti1b on their binding of wild-type EPNR ENTH–Myc. **d, e**, The effect of point (**d**) and helix-deletion (**e**) mutations in EPNR ENTH–Myc on their binding to wild-type GST–Vti1b. **f**, The effect of the charge-swap mutations EPNR(R146E) and Vti1b(E23R) introduced on the basis of the complex structure.

different cargoes and when attached by means of short linkers to the transmembrane segments of reporter constructs.

The specificity of these interactions results from two or three side chains of the motif fitting in pockets on the surface of the domain, with additional strength coming from backbone hydrogen-bond interactions. In EPNR–Vti1b, the mode of interaction is of a different nature, involving the association of two complementary surfaces, the shape and structure of which are dependent on the tertiary structure of both components' domains, with the specificity resulting from a considerable number of weak interactions involving many residues scattered across each half of the interface. The high degree of specificity that this generates is demonstrated by the fact that the SNARE H_{abc} domains from syntaxins 6, 7, 8 and 10 are unable to bind EPNR (Supplementary Fig. 6) and that the other mammalian ENTH-domain-containing proteins, the plasma membrane epsins, are unable to bind Vti1b (Supplementary Fig. 2). The trafficking of all the components of a SNARE complex could be regulated by a single clathrin adaptor if one SNARE from the complex could simultaneously bind both its SNAREpin partners and the adaptor. Such a scenario is demonstrated for Vti1b in Fig. 3a, b, where it is shown that GST–Vti1b in a SNAREpin complex with

syntaxin 7, syntaxin 8 and VAMP8 can bind to EPNR as well as GST–Vti1b can when alone.

To examine the importance of the interaction *in vivo*, stable HeLa cell lines expressing wild-type, Vti1b(E12A) and Vti1b(F73S) Myc-tagged forms of Vti1b were made, and their localization was compared with that of endogenous Vti1b. As seen in Fig. 3c, both Vti1b(E12A) and Vti1b(F73S) were more peripheral and punctate than wild-type Vti1b and endogenous Vti1b, which showed a more juxtanuclear distribution. The change in Vti1b localization resembles that seen when EPNR is depleted by short interfering RNA (Fig. 3d). To quantify the change in distribution of Vti1b, we used an automated microscope to take random photographs of cell lines expressing wild-type and mutant constructs that had been double-labelled for the Myc tag and TGN46. The relative amount of Myc labelling in the TGN46-positive region of the cell was compared with the Myc labelling in the rest of the cell, and showed that the overlap of the two mutant constructs with TGN46 is ~60% that of the wild-type construct (Fig. 3 and Supplementary Fig. 8).

To identify the compartments containing either wild-type or mutant Vti1b Myc-tagged constructs, a number of double-labelling experiments were carried out. Wild-type Vti1b showed partial

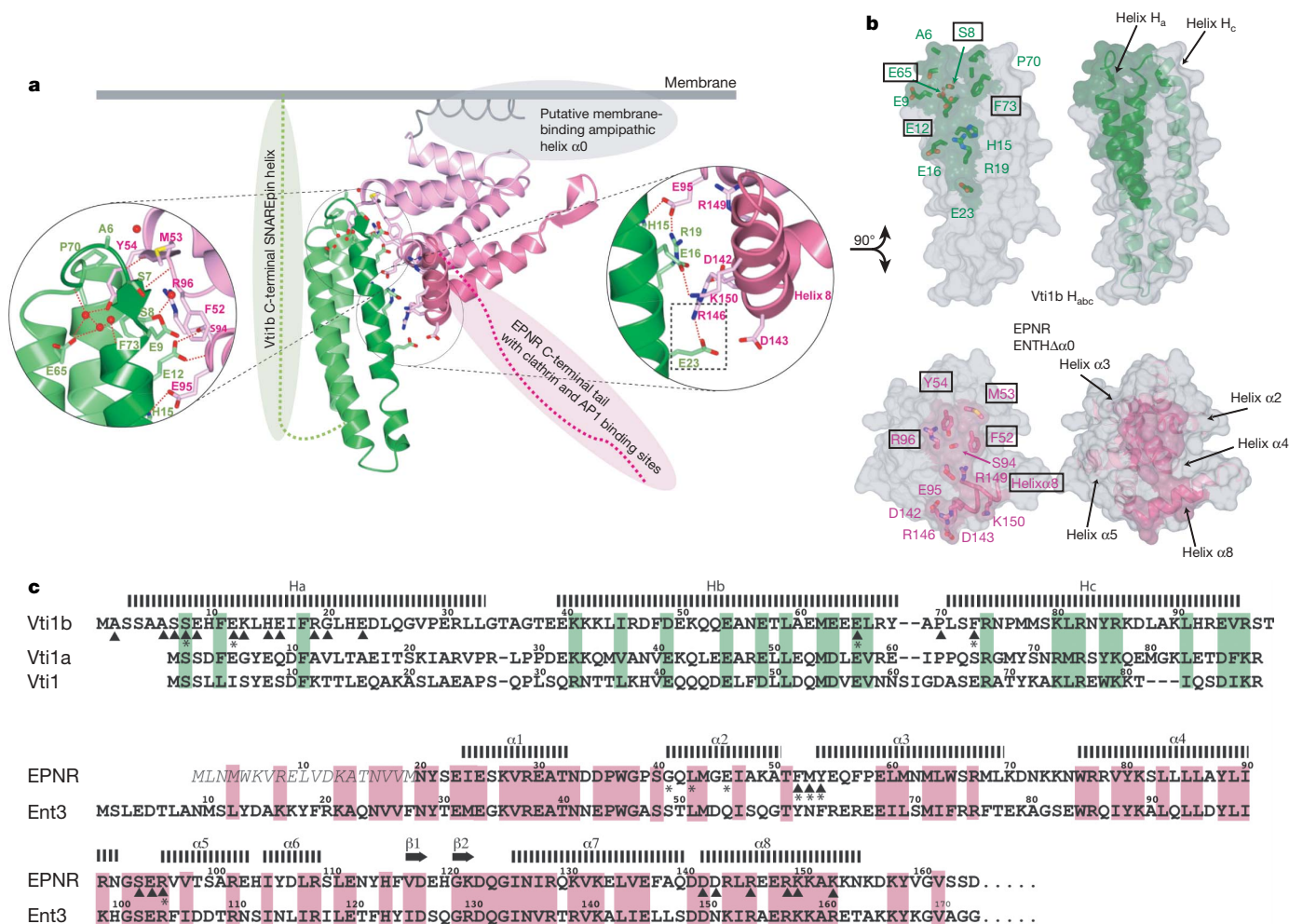


Figure 2 | The EPNR ENTH $\Delta\alpha 0$ –Vti1b H_{abc} domain complex. **a**, The complex is shown, with Vti1b coloured dark green to pale green (N to C) and EPNR coloured pale pink to dark pink (N to C). Enlarged views are shown of key residues in the binding interface. The putative lipid binding helix $\alpha 0$ of EPNR is shown in grey. The proposed orientation of the remaining portions of Vti1b and EPNR are indicated by dotted lines. Charge-swap mutations are boxed. **b**, Surface representations of the complex, with each rotated by 90 degrees to show the ‘footprint’ of interaction coloured green on Vti1b H_{abc} and pink on EPNR ENTH. α helices and side chains participating in the interaction are

shown through the different surfaces. Mutated residues that have been demonstrated to affect binding are boxed. **c**, Structure-based sequence alignment of the H_{abc} domains of mouse Vti1b, mouse Vti1a and yeast Vti1 (with conserved residues boxed in green), and of human EPNR ENTH domain with yeast Ent3 (with conserved residues shown in pink). Residues in which mutation abolishes binding between Vti1b and EPNR are marked with an asterisk. Residues that have significant roles in the binding interface are indicated by a triangle.

colocalization with TGN46 (TGN), EEA1 (early endosome), transferrin (recycling endosome) and LAMP-1 (late endosomes and lysosomes). In contrast, both of the mutant constructs colocalize strongly with LAMP-1-positive late compartments and only weakly with other markers (Fig. 4 and Supplementary Fig. 9a, b). Together, these studies show that the direct Vti1b–EPNR interaction described here is essential for at least one post-Golgi trafficking event undertaken by Vti1b. The decrease in overlap of the two EPNR-binding mutants with TGN46, combined with the increased accumulation of Vti1b in LAMP-1-positive late endosomes and lysosomes, indicates that normally EPNR facilitates retrograde transport of Vti1b from a relatively late endosomal compartment back to an earlier compartment and/or the TGN.

This study provides mechanistic insights into how SNARE proteins can be specifically sorted into CCVs, and in so doing identifies newly discovered functions for both H_{abc} and ENTH domains. In the case of the former, the SNARE H_{abc} domain, which has been implicated in regulating SNARE complex formation and/or tethering¹, has been shown to be essential for directing SNARE trafficking. In the latter case, the phosphoinositide-binding ENTH domain, previously implicated in intracellular membrane localization and membrane

deformation¹⁷, acquires the function of a protein cargo selection/recognition module. The surface–surface association between EPNR and Vti1b must be highly specific to these two components because the topography of the surfaces is dependent on the overall fold of each of the domains, and therefore cannot be readily transplanted onto different combinations of cargoes and cargo adaptors. This surface–surface recognition provides a gain in selectivity at the expense of losing the adaptability and plasticity demonstrated by the widely used linear peptide motif system. Despite the interaction identified here being mechanistically different from that of short linear motifs binding to cargo adaptors—the paradigm of a moderate affinity, dynamic (low micromolar dissociation constant) interaction is maintained because these properties are essential both for controlling the ‘proof-reading’ of cargo incorporated into CCVs and for the dissociation of cargo and adaptors following clathrin-cage disassembly (reviewed in refs 4, 25 and 26). In both cases, however, the effective binding between full-length proteins *in situ* in a forming clathrin coat will be significantly stronger than the binary interactions measured in solution due to avidity effects and coincidence detection^{22,27}. In the case of EPNR–Vti1b, the most prominent additional contributions will come from the membrane itself, because Vti1b is a transmembrane protein and EPNR is expected to bind to membranes by means of helix $\alpha 0$ and will also be held near the membrane through its binding to polymeric clathrin and to AP1 and GGA proteins.

A highly specific non-transplantable interaction is ideally suited to SNARE trafficking, where it is necessary to ensure that sufficient amounts of only the correct SNAREs traverse a direct path between two specific membranes, because by working in parallel with but not competing with the incorporation of standard motif-containing

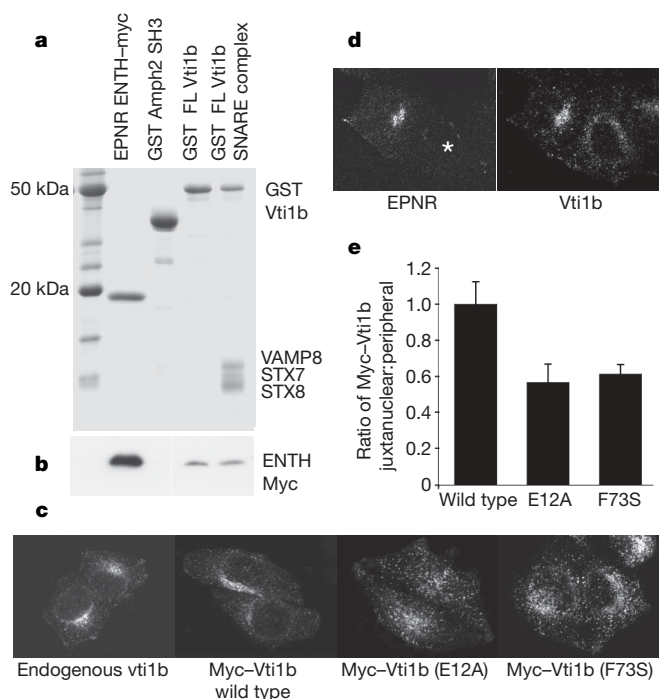


Figure 3 | The effect of disrupting the Vti1b–EPNR interaction on the localization of Vti1b *in vivo*. **a, b**, Full-length (FL) GST Vti1b binds equally well to EPNR when on its own and when in a SNARE complex with syntaxin 7 (STX7), syntaxin 8 (STX8) and VAMP8, whereas a control protein amphiphysin 2 (Amph2) SH3 shows no interaction. **a, b**, Coomassie-blue-stained gel (**a**) and western blot (**b**) of ‘GST pull-down’. **c**, Localization of wild-type and mutant Myc-tagged Vti1b in stably transfected HeLa cells. Clonal cell lines expressing wild-type, E12A and F73S constructs show that the steady-state distribution of Myc–Vti1b wild-type does not differ in juxtanuclear localization from endogenous Vti1b; however, Myc–Vti1b(E12A) and Myc–Vti1b(F73S) both adopt a more peripheral distribution. **d**, Effect of EPNR depletion by siRNA on localization of endogenous Vti1b. Small interfering RNA-treated HeLa cells (asterisk) were mixed with untreated HeLa cells, and were double-labelled for EPNR and Vti1b. **e**, Cells expressing Myc-tagged wild-type and mutant Vti1b constructs were analysed with an automated microscope. The TGN/juxtanuclear region was defined by labelling with anti-TGN46. Fluorescence intensity of the Myc label was quantified in this region and in the rest of the cell. The ratios of juxtanuclear to peripheral Myc labelling displayed by F73S and E12A constructs relative to the wild-type construct are shown. Data from 600 micrographs were used in the analysis; the error bars show the s.d.

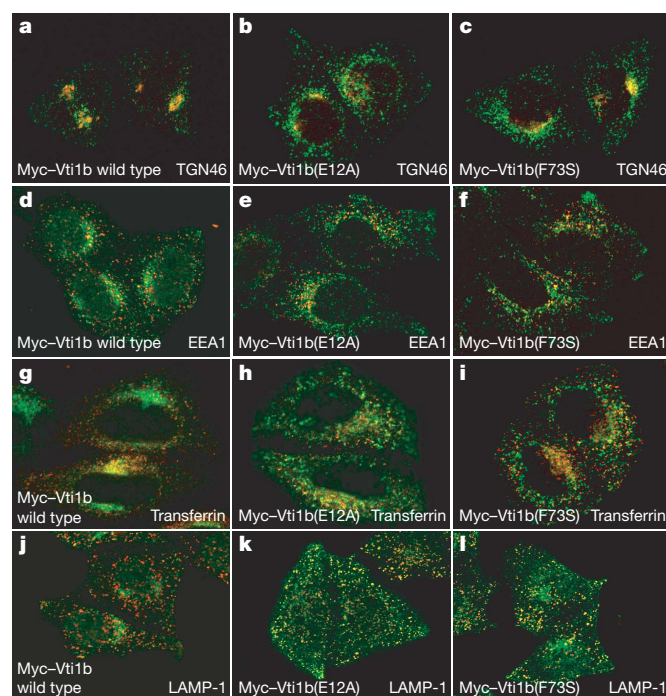


Figure 4 | Disrupting the EPNR–Vti1b interaction *in vivo* increases the amount of Vti1b in late endosomes/lysosomes. Images of stably transfected HeLa cell lines, with Myc-tagged Vti1b constructs in green and various compartment markers in red. All cell lines expressed approximately equal amounts of Myc-tagged Vti1b (see Supplementary Information). At steady state, wild-type Myc–Vti1b shows partial colocalization with markers for all post-Golgi compartments, with the best overlap in the juxtanuclear region and only weak colocalization with LAMP-1-positive late endosome/lysosomes (**a, d, g** and **j**). In contrast, both EPNR non-binding mutants E12A (**b, e, h** and **k**) and F73S (**c, f, i** and **l**) show excellent colocalization with LAMP-1-positive late endosome/lysosomes. Tfr, transferrin.

cargo it ensures that the SNAREs are incorporated, but that the carrying capacity for standard motif-containing cargo is not reduced. In the light of several other non-linear motif SNARE interactions being identified with both CCV coat components (refs 10–12 and also very recently with COPII binding to Sec22b longin domain (ref. 28)), it seems probable that similarly highly specific but mechanistically very different interactions between SNAREs and coat proteins mediated by relatively large numbers of residues on both components, rather than by a short motif on the cargo, will be the paradigm for controlling post-Golgi SNARE transport. The evolution of highly cargo-specific mechanisms for directing the incorporation of SNAREs as cargo into CCVs, as described here, also indicates that analogous highly selective interactions may exist for some non-SNARE cargo to allow their transport along equally specific, restricted routes.

METHODS SUMMARY

Proteins for crystallization were produced as GST fusion proteins from which the GST was cleaved with bovine thrombin and further purified by S200 gel filtration. Vti1b H_{abc} (58 mg ml⁻¹) was crystallized in space group P2₁ against a reservoir containing 100 mM Tris (pH 7), 30% ethanol and 20% glycerol. The structure was solved at a resolution of 2.0 Å by selenomethionine (SeMet) single isomorphous replacement and was refined to final *R*_{cryst} and *R*_{free} values of 23.8% and 30.7%, respectively. EPNR ENTHΔα0 (19 mg ml⁻¹) was crystallized in space group P2₁ against a reservoir containing 30% (w/v) PEG 3000 and 100 mM 2-(N-cyclohexylamino)ethanesulphonic acid (CHES, pH 9.5). The structure was solved at 2.0 Å by molecular replacement using Epn1 ENTH domain as a search model and was refined to final *R*_{cryst} and *R*_{free} values of 18.8% and 25.0%, respectively. Crystals of Vti1b H_{abc}–EPNR ENTHΔα0 complex (space group C2) were grown by mixing the two purified proteins to final concentrations of 4 mM each in a molar ratio of 1:1. Crystals grew over a period of 2–4 months against a reservoir containing 50% (w/v) PEG 200, 100 mM sodium phosphate-citrate (pH 4.2) and 200 mM NaCl, and required no cryo-protection. The structure was solved at 2.2 Å by molecular replacement searching first with ENTHΔα0 and then with Vti1b H_{abc}, and was refined to final *R*_{cryst} and *R*_{free} values of 19.7% and 28.3%, respectively.

Clonal cell lines stably expressing Vti1b were established using pIRESneo2 vectors encoding full-length wild-type and mutant proteins. Immunofluorescence was visualized using antibodies (see Methods). Quantification of Vti1b construct phenotypes were carried out using a Cellomics ArrayScan VTI automated microscope.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 July; accepted 3 October 2007.

- Hong, W. SNAREs and traffic. *Biochim. Biophys. Acta* **1744**, 493–517 (2005).
- Fasshauer, D., Sutton, R. B., Brunger, A. T. & Jahn, R. Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs. *Proc. Natl Acad. Sci. USA* **95**, 15781–15786 (1998).
- Chen, Y. A. & Scheller, R. H. SNARE-mediated membrane fusion. *Nature Rev. Mol. Cell Biol.* **2**, 98–106 (2001).
- Bonifacio, J. S. & Traub, L. M. Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu. Rev. Biochem.* **72**, 395–447 (2003).
- Hurley, J. H., Lee, S. & Prag, G. Ubiquitin-binding domains. *Biochem. J.* **399**, 361–372 (2006).
- Peden, A. A., Park, G. Y. & Scheller, R. H. The Di-leucine motif of vesicle-associated membrane protein 4 is required for its localization and AP-1 binding. *J. Biol. Chem.* **276**, 49183–49187 (2001).
- Tran, T. H., Zeng, Q. & Hong, W. VAMP4 cycles from the cell surface to the trans-Golgi network via sorting and recycling endosomes. *J. Cell Sci.* **120**, 1028–1041 (2007).

- Hirst, J., Miller, S. E., Taylor, M. J., von Mollard, G. F. & Robinson, M. S. EpsinR is an adaptor for the SNARE protein Vti1b. *Mol. Biol. Cell* **15**, 5593–5602 (2004).
- Chidambaram, S., Mullers, N., Wiederhold, K., Haucke, V. & von Mollard, G. F. Specific interaction between SNAREs and epsin N-terminal homology (ENTH) domains of epsin-related proteins in trans-Golgi network to endosome transport. *J. Biol. Chem.* **279**, 4175–4179 (2004).
- Nonet, M. L. et al. UNC-11, a *Caenorhabditis elegans* AP180 homologue, regulates the size and protein composition of synaptic vesicles. *Mol. Biol. Cell* **10**, 2343–2360 (1999).
- Drees, B. L. et al. A protein interaction map for cell polarity development. *J. Cell Biol.* **154**, 549–571 (2001).
- Martinez-Arca, S. et al. A dual mechanism controlling the localization and function of exocytic v-SNAREs. *Proc. Natl Acad. Sci. USA* **100**, 9011–9016 (2003).
- Pryor, P. R. et al. Combinatorial SNARE complexes with VAMP7 or VAMP8 define different late endocytic fusion events. *EMBO Rep.* **5**, 590–595 (2004).
- Murray, R. Z., Wylie, F. G., Khromykh, T., Hume, D. A. & Stow, J. L. Syntaxin 6 and Vti1b form a novel SNARE complex, which is up-regulated in activated macrophages to facilitate exocytosis of tumor necrosis factor-α. *J. Biol. Chem.* **280**, 10478–10483 (2005).
- Song, J., Lee, M. H., Lee, G. J., Yoo, C. M. & Hwang, I. *Arabidopsis* EPSIN1 plays an important role in vacuolar trafficking of soluble cargo proteins in plant cells via interactions with clathrin, AP-1, VTI11, and VSR1. *Plant Cell* **18**, 2258–2274 (2006).
- Mossessova, E., Bickford, L. C. & Goldberg, J. SNARE selectivity of the COPII coat. *Cell* **114**, 483–495 (2003).
- Ford, M. G. et al. Curvature of clathrin-coated pits driven by epsin. *Nature* **419**, 361–366 (2002).
- Krissinel, E. & Henrick, K. *Detection of Protein Assemblies in Crystals* (ed. Berthold, M.) (Springer, Berlin/Heidelberg, 2005).
- Fridmann-Sirkis, Y., Kent, H. M., Lewis, M. J., Evans, P. R. & Pelham, H. R. Structural analysis of the interaction between the SNARE Tlg1 and Vps51. *Traffic* **7**, 182–190 (2006).
- Munson, M., Chen, X., Cocina, A. E., Schultz, S. M. & Hughson, F. M. Interactions within the yeast t-SNARE Sso1p that control SNARE complex assembly. *Nature Struct. Biol.* **7**, 894–902 (2000).
- Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
- Honing, S. et al. Phosphatidylinositol-(4,5)-bisphosphate regulates sorting signal recognition by the clathrin-associated adaptor complex AP2. *Mol. Cell* **18**, 519–531 (2005).
- Owen, D. J. & Evans, P. R. A structural explanation for the recognition of tyrosine-based endocytotic signals. *Science* **282**, 1327–1332 (1998).
- Kato, Y., Misra, S., Puertollano, R., Hurley, J. H. & Bonifacio, J. S. Phosphoregulation of sorting signal–VHS domain interactions by a direct electrostatic mechanism. *Nature Struct. Biol.* **9**, 532–536 (2002).
- Edeling, M. A., Smith, C. & Owen, D. Life of a clathrin coat: insights from clathrin and AP structures. *Nature Rev. Mol. Cell Biol.* **7**, 32–44 (2006).
- Traub, L. M. Common principles in clathrin-mediated sorting at the Golgi and the plasma membrane. *Biochim. Biophys. Acta* **1744**, 415–437 (2005).
- Di Paolo, G. & De Camilli, P. Phosphoinositides in cell regulation and membrane dynamics. *Nature* **443**, 651–657 (2006).
- Mancias, J. D. & Goldberg, J. The transport signal on Sec22 for packaging into COPII-coated vesicles is a conformational epitope. *Mol. Cell* **26**, 403–414 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank PX beamline staff at SRS Daresbury and ESRF ID23-1, J. Connell for assistance with microscopy, and P. Luzio and P. Evans for discussions. This work was supported by a Wellcome Trust SRF to D.J.O., a PRF to M.S.R., an MRC studentship to S.E.M. and an Australian NHMRC Career Development Award to B.M.C.

Author Information Coordinates have been deposited under PDB codes 2QYW, 2QY7 and 2V8S for Vti1b H_{abc}, epsin ENTH domain and the Vti1b/epsinR complex, respectively. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.J.O. (djo30@cam.ac.uk) and M.S.R. (msr12@mole.bio.cam.ac.uk).

METHODS

Plasmid construction. The DNA encoding the human EPNR ENTH domain (residues 1–166) was cloned with a C-terminal Myc tag into pMW172H6 for expression with an N-terminal His tag and a C-terminal Myc tag (giving His₆-EPNR-ENTH-Myc). Residues 20–166 were cloned into pGEX-4T-2 giving GST EPNR ENTHΔα0. Residues 20–166 and 1–142 were cloned into pMWH6Myc to give H₆-EPNR-ENTHΔα0-Myc and His₆-EPNR-ENTHΔα8-Myc, respectively.

DNA encoding rat Epn1 ENTH (residues 1–163) was cloned into pMWH6Myc such that both tags were at the N terminus of the domain to give His₆-Myc-Epn1-ENTH. Residues 1–95 of rat synaptobrevin were cloned into pMWGST such that the GST was in-frame following the synaptobrevin fragment to produce synaptobrevin-GST.

DNA encoding mouse Vti1b lacking the transmembrane domain (residues 1–206) was cloned into pGEX-4T-2 for expression with an N-terminal GST tag, as were three shorter constructs: Vti1b(1–96) (Vti1b H_{abc}), Vti1b(1–116) and Vti1b(117–206). Wild-type, E12A and F73S mutant N-terminal Myc-tagged full-length Vti1b constructs were cloned into pIRESneo2 vector (Clontech, BD Biosciences) and used to generate stably transfected cell lines. DNA encoding the yeast Ent3 ENTH domain (residues 1–174) and Vti1 (both full-length and truncated) were cloned into pMW172H6 and pGEX-4T-2, respectively. All point mutations were generated using the Quikchange mutagenesis kit (Stratagene).

The SNARE helices of human syntaxin 7 (169–237), human syntaxin 8 (149–213) and VAMP (1–76) were cloned into pMW172H6 for expression with an N-terminal His tag. The H_{abc} domains of human syntaxin 6 (1–173), syntaxin 7 (1–175), syntaxin 8 (1–155) and syntaxin 10 (1–167) were cloned into pGEX-4T2 (see Supplementary Fig. 1).

Protein binding assays. All proteins were expressed in *E. coli* BL21(DE3)/pLysS at 22 °C for 16 h after induction with 0.2 mM isopropyl-β-D-thiogalactoside on reduction of temperature from 37 °C. GST fusion proteins were purified on glutathione-sepharose, and His₆-tagged proteins were purified on Ni-NTA agarose, by standard procedures. For standard GST pull-down experiments, 2 nmol GST fusion protein was incubated in 1 ml PBS supplemented with 0.1% NP-40 (buffer C) and 2 nmol of the required His₆/Myc-tagged protein at 20 °C for 30 min before 120 μl of 50% slurry of glutathione-sepharose was added. This was then incubated for 30 min and the beads pelleted by centrifugation. Pellets were washed three times with 1 ml buffer C, and bound proteins were eluted by boiling in SDS sample buffer. Bound proteins were analysed by SDS-PAGE, and were stained with Coomassie blue or probed by western blotting with monoclonal anti-Myc (9E10, Santa Cruz) followed by a rabbit anti-mouse secondary antibody and then ¹²⁵I-labelled protein A.

For SNARE formation, 2 nmol of the full-length cytosolic domain of GST-Vti1b was incubated with 10 nmol of each of the His-tagged SNARE helix domains of human syntaxin 6, syntaxin 7 and VAMP8 for 48 h in 10 mM Tris (pH 8.7), 200 mM NaCl and 1 mM DTT (buffer A) in the presence of 60 μl of 50% slurry of glutathione-sepharose. The beads were pelleted by centrifugation and were washed three times with 1 ml buffer A. Stoichiometric SNARE-complex formation and binding of the complex to the beads was shown, by Coomassie blue staining of SDS-PAGE, to occur only in the presence of GST-Vti1b (data not shown). The ability of this complex to bind EPNR ENTHΔα0 was compared with that of 2 nmol of uncomplexed full-length GST-Vti1b cytosolic domain and 2 nmol of a control GST protein (GST-amphiphysin 2) SH3 domain, as in a standard pull-down experiment (see above).

Isothermal titration calorimetry. For ITC, each protein was buffer-exchanged into 50 mM HEPES pH 7.4 and 300 mM NaCl by gel filtration on Superdex S200. A VP-ITC instrument (MicroCal Inc.) was used at 20 °C. Typically the ITC cell contained 1.4 ml EPNR ENTH-Myc protein at a concentration of 180 μM. Vti1b was added in 30–40 injections of 4–8 μl at a concentration of 1.7 mM. The heat of dilution of the Vti1b protein into buffer was subtracted from the data, and titration curves were fitted using ORIGIN, yielding values for the stoichiometry *n*, equilibrium association constant *K*_a (= *K*_d⁻¹) and enthalpy of binding.

Protein purification for structure determination. GST-tagged Vti1b(1–96) was expressed in BL21(DE3)/pLysS cells for 16 h at 22 °C after induction with 0.2 mM isopropyl-β-D-thiogalactoside. Cells were lysed in solution containing 10 mM Tris (pH 8.7), 200 mM NaCl, 1 mM DTT (buffer A) and protease inhibitors, and the protein was affinity-purified on glutathione-sepharose. The GST tag was cleaved from Vti1b(1–96) with thrombin overnight at 20 °C while the protein was bound to the beads, and the cleaved protein was eluted. The protein was further purified by gel-filtration in buffer A, and concentrated to 58 mg ml⁻¹. To incorporate SeMet, the cells were grown in M9 minimal media in the presence of 50 μg ml⁻¹ L-SeMet²⁹. SeMet protein was purified similarly to native protein in buffer A supplemented with 10 mM DTT. The best crystals of both native and SeMet crystals were grown by sitting-drop vapour diffusion against a reservoir containing 100 mM Tris (pH 7), 30% ethanol and 20%

glycerol. Crystals (space group *P*₂₁) grew in 2–3 days and required no further cryoprotection.

Initial attempts to purify and determine the structure of the whole EPNR ENTH domain (residues 1–166) were unsuccessful because the protein was poorly soluble and prone to aggregation. Deletion of residues 1–19 including the putative amphipathic helix α0, which by analogy to helix α0 in Epn1 may partially embed itself in phospholipid membranes^{17,30}, resulted in a readily soluble protein, termed ENTHΔα0, the circular dichroism spectra of which showed no significant difference from the full-length ENTH domain (data not shown). These circular dichroism data indicate that the putative helix α0 does not adopt a helical conformation in solution in the absence of a phospholipid membrane. N-terminal GST-tagged EPNR (20–166) (EPNR ENTHΔα0) was expressed and purified similarly to GST-Vti1b(1–96), except that buffer B (10 mM HEPES (pH 7.4) and 300 mM NaCl) was used throughout. The best crystals were grown over the space of several weeks by sitting-drop vapour diffusion from 19 mg ml⁻¹ protein against a reservoir containing 30% (w/v) PEG 3000 and 100 mM CHES (pH 9.5). Crystals (space group *P*₂₁) were cryoprotected in reservoir buffer containing 20% glycerol.

Crystals of Vti1b H_{abc}-EPNR ENTHΔα0 complex (space group *C*₂) were grown by mixing the two individually purified proteins to final concentrations of 4 mM each in a molar ratio of 1:1 after the exchange of Vti1b into buffer B by gel filtration. The best crystals grew over a period of 2–4 months by sitting-drop vapour diffusion against a reservoir containing 50% (w/v) PEG 200, 100 mM sodium phosphate-citrate (pH 4.2) and 200 mM NaCl, and required no further cryoprotection.

Crystallographic structure determinations. The Vti1b H_{abc} domain crystallized in the spacegroup *P*₂₁ with unit cell dimensions *a* = 24.23 Å, *b* = 25.17 Å, *c* = 80.12 Å and angles α = 90.00°, β = 95.23°, γ = 90.00°, and diffracted to a resolution of 2.0 Å. X-ray data of native Vti1b were collected in-house at 100 K with a Rigaku rotating anode X-ray source fitted with a MAR345 image plate detector. Data for SeMet Vti1b protein crystals were collected using the Daresbury SRS beamline 9.6 at 100 K equipped with a CCD detector. X-ray diffraction data from both the native and SeMet crystals were integrated using MOSFLM³¹ and were scaled with SCALA³². The selenium sites were determined from anomalous Pattersons using AUTOSHARP³³. Heavy-atom refinement and phasing was done with SHARP³⁴ using the method of single isomorphous replacement with anomalous scattering (SIRAS) incorporating the selenium and native data, followed by solvent flipping and flattening with SOLOMON. Model building was carried out in O³⁵, and the structure refined with REFMAC5³⁶ with data to 2.0 Å. All residues including the loop regions had visible electron density. Final *R*_{cryst} and *R*_{free} values were obtained of 22.5% and 29.6%, respectively, and the model contains no Ramachandran violations (Supplementary Fig. 5). We expect that the relatively high *R*-factors are owing to a large degree of disorder in loop regions, which constitute a significant fraction of the scattering mass. MLHL refinement was attempted, but did not result in major improvements.

The EPNR ENTHΔα0 domain crystallized in the spacegroup *P*₂₁ with unit cell dimensions *a* = 46.5 Å, *b* = 91.1 Å, *c* = 53.0 Å and angles α = 90.0°, β = 107.0°, γ = 90.00°, with three molecules in the asymmetric unit (ASU), and diffracted to a resolution of 2.0 Å. X-ray data of EPNR ENTHΔα0 were collected at ESRF ID23-1 at 100 K on a CCD detector, and integrated and scaled as for Vti1b. EPNR ENTHΔα0 was solved by molecular replacement with a search model of the Epn1 ENTH domain (residues 20–160) using PHASER³⁷. The structure was built into the initial electron density map, and was then subjected to alternating rounds of rebuilding in COOT³⁸ and refinement with REFMAC5³⁶. There were 135 residues with visible electron density. Final *R*_{cryst} and *R*_{free} values were obtained of 18.8% and 25.0%, respectively (Supplementary Fig. 5), and the model contained no Ramachandran violations.

The EPNR ENTHΔα0-Vti1b H_{abc} complex crystallized in the spacegroup *C*₂ with unit cell dimensions *a* = 102.2 Å, *b* = 60.9 Å, *c* = 41.0 Å and angles α = 90.0°, β = 94.0°, γ = 90.0°, with one molecule in the ASU, and diffracted to a resolution of 2.2 Å. X-ray data of EPNR ENTHΔα0-Vti1b H_{abc} were collected in-house at 100 K with a Rigaku rotating anode X-ray source fitted with a MAR345 image plate detector. X-ray diffraction data were integrated, and were scaled as for Vti1b. The complex was solved, again using PHASER³⁷, searching first with the EPNR ENTHΔα0 domain structure and subsequently, once this had been positioned, searching with and positioning the Vti1b H_{abc} structure. Building and refinement was carried out as for EPNR ENTHΔα0 domain alone, and the resulting final structure had an *R*_{cryst} of 19.7% and an *R*_{free} 28.3% (Supplementary Fig. 5). The model contained no Ramachandran violations. Statistics for all structure determinations are given in the Supplementary Information.

Transfection and immunolocalization. pIRESneo2 vectors encoding full-length (residues 1–232) wild-type Vti1b and F73S and E12A mutants were

transfected into HeLa cells using Eugene 6 (Roche). Stably transfected cells were selected with G418 (Geneticin, Invitrogen), from which clonal lines were established by culturing isolated colonies under continuous G418 selection. For immunofluorescence, cells were fixed with methanol or PFA, and transfected proteins and compartment markers were visualized using the following antibodies: mouse monoclonal anti-Vti1b (BD Transduction Laboratories), monoclonal anti-Myc (9E10, Santa Cruz Biotechnology), rabbit polyclonal anti-Myc (Cell Signaling Technologies), rabbit polyclonal anti-EPNR, monoclonal anti-EEA1 (Serotech), monoclonal anti-LAMP-1 (gift of P. Luzio), sheep polyclonal anti-TGN46 (BD Transduction Laboratories) or with Alexa Fluor 555 transferrin.

A short interfering RNA (siRNA) oligonucleotide (Dharmacon) targeted against EPNR (AAGAAGAGUUUAUAAGUCGUU) was used to deplete EPNR. The oligofectamine reagent (50 μ l) was added to 100 μ l Opti-MEM (both Invitrogen) and was incubated for 5 min. EPNR siRNA oligo (25 μ l from 40 μ M stock) was diluted with 825 μ l Opti-MEM, was added to Opti-MEM/oligofectamine mix and was incubated at room temperature for 15 min; thereafter, 4 ml Opti-MEM was added. Fifty-per-cent confluency cells in a 9-cm² dish were washed with Opti-MEM after DMEM had been removed. The 5 ml of oligonucleotide-oligofectamine mix was added and incubated for 4 h at 37 °C. The transfection was inhibited by the addition of 5 ml DMEM, supplemented with 20% FCS. After 12–14 h, fresh DMEM was added. After a further 8 h the siRNA-treated cells were trypsinized and mixed at a 1:1 ratio with standard HeLa cells, and were cultured onto coverslips for immunofluorescence microscopy.

For quantitative microscopy, cell lines expressing Myc-tagged wild-type, E12A and F73S Vti1b constructs were grown in 96-well plates. Two cell lines were used for each of the three constructs, and each line was split into 20 wells, making a total of 120 wells. The cells were fixed with paraformaldehyde, permeabilized with methanol, and double-labelled with monoclonal anti-Myc (9E10, Santa Cruz Biotechnology) and sheep polyclonal anti-TGN46 (BD Transduction Laboratories). Images from five fields were acquired for each well, using a Cellomics ArrayScan VTI automated microscope and focusing on the TGN46 stain. Image analysis was performed using the Cellomics Molecular Translocation Bioapplication. The TGN46-positive ('circ') region was identified using an intensity threshold, then the fluorescence intensity of the Myc label was measured both in this Circ region and in the outlying ('ring') region. The ratio of Myc labelling in the two areas (circ:ring average intensity ratio) was quantified for each of the 120 wells, using 5 images for each well. These ratios were then used to calculate the mean ratio and standard deviation for each of the three constructs.

29. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124 (1993).
30. Stahelin, R. V. *et al.* Contrasting membrane interaction mechanisms of AP180 N-terminal homology (ANTH) and epsin N-terminal homology (ENTH) domains. *J. Biol. Chem.* **278**, 28993–28999 (2003).
31. Leslie, A. G. W. in *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography* No. 26 (SERC, Daresbury Laboratory, Warrington, 1992).
32. Evans, P. R. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
33. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2006).
34. de la Fortelle, E. & Bricogne, G. in *Methods in Enzymology* (eds Carter, C. W. Jr & Sweet, R. M.) 472–494 (1997).
35. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
36. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
37. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., Adams, P. D. & Read, R. J. PHASER crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
38. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

Identification of a mechanism of photoprotective energy dissipation in higher plants

Alexander V. Ruban^{1*}, Rudi Berera^{2*}, Cristian Illoiaia^{3,4}, Ivo H. M. van Stokkum², John T. M. Kennis², Andrew A. Pascal³, Herbert van Amerongen⁵, Bruno Robert³, Peter Horton⁴ & Rienk van Grondelle²

Under conditions of excess sunlight the efficient light-harvesting antenna¹ found in the chloroplast membranes of plants is rapidly and reversibly switched into a photoprotected quenched state in which potentially harmful absorbed energy is dissipated as heat^{2,3}, a process measured as the non-photochemical quenching of chlorophyll fluorescence or qE. Although the biological significance of qE is established^{4–6}, the molecular mechanisms involved are not^{7–9}. LHCII, the main light-harvesting complex, has an inbuilt capability to undergo transformation into a dissipative state by conformational change¹⁰ and it was suggested that this provides a molecular basis for qE, but it is not known if such events occur *in vivo* or how energy is dissipated in this state. The transition into the dissipative state is associated with a twist in the configuration of the LHCII-bound carotenoid neoxanthin, identified using resonance Raman spectroscopy¹¹. Applying this technique to study isolated chloroplasts and whole leaves, we show here that the same change in neoxanthin configuration occurs *in vivo*, to an extent consistent with the magnitude of energy dissipation. Femtosecond transient absorption spectroscopy¹², performed on purified LHCII in the dissipative state, shows that energy is transferred from chlorophyll *a* to a low-lying carotenoid excited state, identified as one of the two luteins (lutein 1) in LHCII. Hence, it is experimentally demonstrated that a change in conformation of LHCII occurs *in vivo*, which opens a channel for energy dissipation by transfer to a bound carotenoid. We suggest that this is the principal mechanism of photoprotection.

The twisted configuration of neoxanthin in quenched LHCII crystals and aggregates is detected by the appearance of a band at 953 cm⁻¹ in the ν_4 region of the Raman spectrum¹¹. Because almost all of the neoxanthin found in thylakoid membranes is bound to LHCII (ref. 13), this information can be used to probe the configuration of neoxanthin in isolated chloroplasts and whole leaves, and hence determine the presence of the quenched conformation of LHCII. *In vivo*, Raman signals arise from all the carotenoids in the thylakoid membrane, but because only neoxanthin is in a 9-*cis* conformation it exhibits fingerprint Raman bands¹⁴, which allow its contribution to be measured; this was 50–65% of the 953 cm⁻¹ signal (Supplementary Information).

Resonance Raman spectra were obtained of isolated *Arabidopsis* chloroplasts taken directly from the light (+qE) and 5 min after transfer to darkness (-qE), which resulted in collapse of the Δ pH and consequent relaxation of qE (Fig. 1a). The Raman spectra in the ν_4 region show two main bands at around 953 cm⁻¹ and 964 cm⁻¹. On qE induction, there is an enhancement of the 953 cm⁻¹ band

relative to the 964 cm⁻¹ band by 5–7%, similar to the change previously observed on an increase in quenching in isolated LHCII (ref. 10). After normalization at 1,003 cm⁻¹ (the ν_3 region of the spectrum), the neoxanthin-specific spectra associated with quenching were calculated (Fig. 1b): the qE-associated spectrum shows three

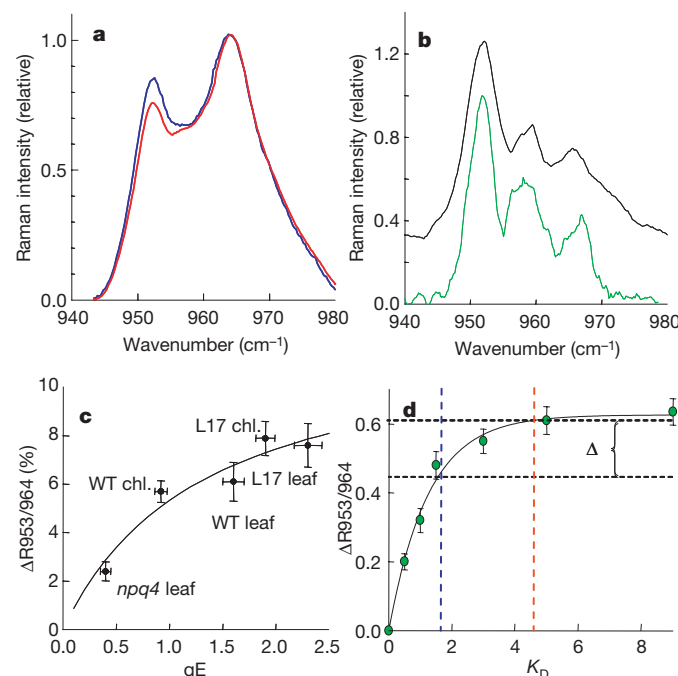


Figure 1 | Quenching-related changes in the neoxanthin region of the resonance Raman spectrum of isolated LHCII, chloroplasts and leaves. **a**, The ν_4 region of the resonance Raman spectra of wild-type chloroplasts after illumination (blue) and following a 5-min relaxation in the dark (red). **b**, Quenched - unquenched difference Raman spectra for qE (green) and for LHCII (black). For qE, the spectra were as in **a**. For LHCII unquenched and quenched, the K_d were 0 and 9, respectively (Supplementary Information). **c**, The extent of qE compared to the change in Raman intensity at 953 cm⁻¹ for leaves and chloroplasts (chl.) from wild-type (WT), L17 and *npq4* plants. The data are means of *n* replicates \pm s.e.m. *n* = 17 (WT chl.), 8 (L17 chl.), 3 (*npq4* leaf), 4 (WT leaf) and 3 (L17 leaf). Twenty-five spectra were recorded for each sample. The line shown is the best fit (95% confidence by *t*-test). **d**, Change in the relative Raman intensity at 953 cm⁻¹ for LHCII with different extents of fluorescence quenching, K_q . For details, see text. Error bars are the calculated amplitudes of noise on the spectra.

¹School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK. ²Department of Physics and Astronomy, Faculty of Sciences, VU University Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands. ³Commissariat à l'Energie Atomique (CEA), Institut de Biologie et Technologies de Saclay (iBiTecS) and Centre National de la Recherche Scientifique (CNRS), Gif-sur-Yvette, F-91191, France. ⁴Department of Molecular Biology and Biotechnology, University of Sheffield, Western Bank, Sheffield S10 2TN, UK. ⁵Laboratory of Biophysics, Wageningen University, PO Box 8128, 6700 ET, Wageningen, The Netherlands.

*These authors have contributed equally to this work.

characteristic bands, which were also found in the quenched spectrum for LHCII. We then obtained further Raman spectra for both leaves and chloroplasts, not only from wild-type *Arabidopsis*, but also from the *npq4-1* mutant¹⁵, which shows much-reduced qE, and from the PsbS-overexpressing L17 line, which shows increased qE¹⁶. In all cases, the extent of the enhancement of the 953 cm^{-1} band was correlated with the magnitude of qE (Fig. 1c), the maximum change being approximately 8%. The experimental design ensured that the changes in the Raman spectra were associated with qE, and did not arise from an alteration in the level of zeaxanthin: first, no epoxidation of zeaxanthin occurred during the 5-min dark relaxation period; and second, the *npq4-1* mutant showed the same level of violaxanthin de-epoxidation as the wild type and L17 mutant (data not shown). It is concluded that the formation of qE is associated with a change in configuration of neoxanthin, the same as found in quenched LHCII.

To estimate how much of a change in LHCII-bound neoxanthin would be predicted for a given change in qE, isolated LHCII was obtained in a range of oligomerization states, which cover a tenfold difference in fluorescence yield. There was a progressive increase of the intensity of the 953 cm^{-1} band on LHCII oligomerization (Supplementary Information), which correlated with the extent of fluorescence quenching (K_d) for each LHCII sample (Fig. 1d). A comparison of the *in vivo* spectrum in the unquenched state with the *in vitro* data indicate that LHCII is in a partially quenched state, corresponding to a K_d of 1.5–2.0 (blue dotted line in Fig. 1d). Hence, the change in K_d arising from qE starts from a K_d of 1.5–2.0 and

reaches a maximum of 4.0–4.5 (red dotted line). Thus, we predict a maximum increase in intensity of the 953 cm^{-1} neoxanthin band of about 0.12 (approximately 20%) for an increase in K_d of 2.0 (Δ , between the two horizontal dotted lines). Taking into account the contributions of other xanthophylls (see above), the maximum predicted qE-related change in the intensity of the 953 cm^{-1} band *in vivo* would be around 10–13%—of the same order as the 8% change observed. It is concluded that under conditions of qE, a conformational change in LHCII occurs that gives rise to a significant amount of non-photochemical quenching.

We used femtosecond transient absorption spectroscopy combined with spectrally dispersed detection and global analysis¹² to determine the mechanism of quenching in isolated LHCII (and therefore in qE), specifically testing proposals that the excited chlorophyll states decay via the population of a carotenoid excited and/or charge transfer state^{17–19}. Transient absorption data following a 100 fs laser pulse were obtained for the unquenched and quenched samples of LHCII, which have an approximate K_d of 0 and 9, respectively. The time traces were very different for the two samples (compare Fig. 2a and b). The spectral data at 677 nm show the initial bleaching of the chlorophyll ground state absorption, which relaxes as the excited state decays to the ground state (top). The traces describe a multi-exponential process consisting of three phases: a ~ 1 ps component that results from excitation equilibration; a ~ 20 ps component assigned to singlet–singlet annihilation of chlorophyll excited states; and a slower component of ~ 1 ns (unquenched) and ~ 130 ps (quenched), which corresponds to their different chlorophyll

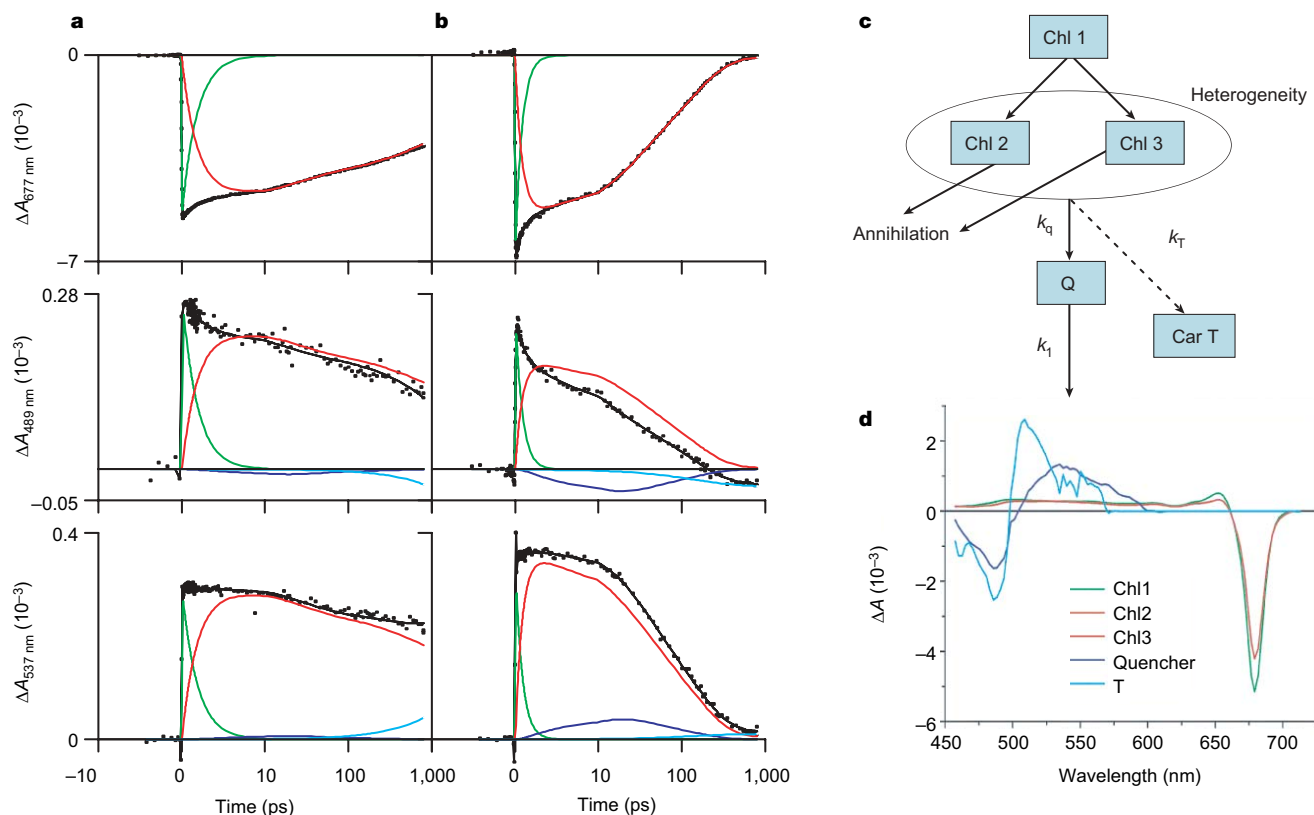


Figure 2 | Femtosecond spectroscopy of LHCII in the unquenched and quenched states. **a, b**, Transient absorption traces for unquenched (**a**) and quenched (**b**) LHCII at 677, 489 and 537 nm. y axis, absorption change, $\Delta A \times 10^{-3}$; x axis, linear from -10 to 10 ps and logarithmic thereafter. The curves (black) were fitted using the target analysis model (**c**, and Supplementary Information). Green, 1 ps phase due to chlorophyll excited-state relaxation; red, chlorophyll excited-state decay; blue, absorption changes due to the quencher Q; cyan, build-up of the triplet state. **c**, The model used to fit the data for quenched LHCII, with 5 compartments: at time

zero the excitation resides in the first compartment (Chl 1), which relaxes on a ps timescale to Chl 2 and Chl 3. Chls 2 and 3 are introduced to account for the fraction of aggregates in which excitations are quenched by singlet–singlet annihilation (on a timescale of several tens or hundreds of ps). Both Chl 2 and Chl 3 are quenched via the quenching state Q with a rate constant k_q , and populate the triplet species Car T, with a rate constant k_T . **d**, Species-associated difference spectra (SADS). Green, the initial excited state (Chl 1); red, Chl 2 and 3; dark blue, the quenching state Q; cyan, Car T.

excited-state lifetimes. The traces at 489 nm (middle) and 537 nm (bottom)—wavelengths that are in the regions of carotenoid ground-state bleach and excited-state absorption, respectively—are only slightly different for the unquenched sample. However, for the quenched sample these traces differ markedly: the 489 nm trace shows decay on the 10–20 ps timescale that is absent in the 537 nm trace. In contrast, the decay at 537 nm is slower on the 10–20 ps timescale when compared to the chlorophyll decay at 677 nm. These differences suggest that in the quenched sample, concomitant with the decay of the chlorophyll excited states, another species is transiently populated, which we will identify as a carotenoid S_1 state—a low-lying optically forbidden excited state. In samples with a K_d of approximately 2, the same carotenoid feature was observed, but to a lesser extent (Supplementary Information). Extensive probing of the 900–1,000 nm region in the same time window did not reveal any absorption changes that could be ascribed to a carotenoid radical (Supplementary Information).

To identify the spectrum of the quenching state, a target analysis was applied to the time-resolved data¹² (Fig. 2c, d). The results from the target analysis demonstrate the essential role of the carotenoid S_1 state in the quenching process. The chlorophyll spectrum shows the bleach of the chlorophyll Q_y state in the 675 nm region, a dip around 615 nm, corresponding to the bleach of the chlorophyll Q_x state, and a region of almost flat excited-state absorption. The cyan spectrum corresponds to the long-lived carotenoid triplet state. The spectrum of the quencher (blue line) reveals the following features: excited state absorption between 505 and 600 nm, corresponding to the carotenoid $S_1 \rightarrow S_n$ transition; and ground-state bleach below 505 nm, corresponding to the bleach of the carotenoid $S_0 \rightarrow S_2$ transition. With the same model applied to the unquenched sample the quenching state remained almost unpopulated. The target analysis yielded an excellent fit to the transient absorption traces (Fig. 2a, b). Most importantly, in quenched LHCII, there is a significant deviation between the chlorophyll excited-state decay kinetics and the traces measured at 489 and 537 nm, owing to population of the carotenoid S_1 state (blue lines); however, in the unquenched sample, these kinetics are the same.

The spectral evolution in the carotenoid absorption region is strongly reminiscent of that observed in artificial carotenoid—phthalocyanine dyads, in which it was demonstrated that quenching

occurred via the population of the carotenoid S_1 state¹⁹. Thus, in the same way, our data provide unequivocal evidence for the population of a carotenoid excited state that is concomitant with the quenching of chlorophyll excited states in aggregates of LHCII, strongly pointing to chlorophyll *a*–carotenoid energy transfer as the quenching mechanism. The samples used here contain only trace levels of violaxanthin (and no zeaxanthin), so lutein or neoxanthin must be responsible for the absorption changes associated with quenching. A comparison of the bleach of the quenching state with the bleach of the carotenoid triplet state shows that their corresponding negative-peak positions coincide. The maximum bleach in the carotenoid triplet spectrum corresponds to lutein 1 (ref. 20). Consequently, lutein 1 is also likely to be the quencher. Lutein 1 is found in an LHCII domain containing chlorophyll *a* 610, chlorophyll *a* 611 and chlorophyll *a* 612 (ref. 21), where the excitation has the highest possibility to be localized, and therefore it is the obvious site for quenching²².

Studies using isolated LHCII and related antenna complexes have provided important insights into the mechanism of qE^{2,23,24}. Most importantly, it was shown that crystallized LHCII is in a quenched state¹⁰, demonstrating that energy dissipation was an intrinsic feature of each LHCII molecule that is brought about by specific configurations of the chlorophylls and xanthophylls bound to the complex, and proving that LHCII can exist in different conformational states that have differing capacities for energy dissipation. However, there was no direct proof that the proposed conformational changes occur *in vivo*; such proof has now been provided from Raman spectroscopy of chloroplasts and leaves in different qE states. The same twisted configuration in neoxanthin was found in both quenched LHCII *in vitro* and in thylakoid membranes *in vivo* under conditions in which the rapidly reversible, Δ pH-dependent (qE) form of non-photochemical quenching is present. We conclude that quenching occurs under these conditions by the same process as *in vitro*. Therefore we re-affirm a model in which qE is initiated by a conformational change in LHCII, which is induced by the light-dependent Δ pH (ref. 2). We propose that this change in conformation gives rise to an increase in the rate of energy transfer to lutein 1 and, consequently, to energy dissipation (Fig. 3). LHCII aggregation has previously been associated with a conformational change in the lutein 1 domain²⁵. Moreover, from a comparison of lutein 1 and lutein 2 in the crystal structure, it has been speculated how a change in the configuration of lutein 1 would bring it closer to chlorophyll *a* 612 (ref. 26), providing the key step in the switching on of quenching. This idea provides an explanation of the link between the observed changes in protein conformation and fluorescence quenching.

Although this model can fully account for qE both qualitatively and quantitatively, qE may be a heterogeneous process of which quenching by lutein 1 in LHCII is only a part. The formation of a radical state of zeaxanthin has been correlated with qE *in vivo*, suggesting that it is either the quencher itself or its formation is closely associated with the quenching process⁷. Although we did not detect the presence of a charge-separated state involving a carotenoid radical in quenched LHCII aggregates, the LHCII samples used here do not contain any zeaxanthin. Similarly, our data do not exclude the possibility that quenching also occurs in the minor antenna complexes, CP24, CP26 and CP29, which have been reported to bind zeaxanthin at sites occupied by lutein in the major LHCII (ref. 24). Further work is needed to establish the relationship between different xanthophyll excited states and radical states *in vivo*, and the contributions they make not only to qE but to other more sustained energy dissipation states that can arise under certain extreme conditions of plant stress²⁷ and that seem to involve formation of LHCII aggregates similar to those used here²⁸. Clearly, however, both the conformation of antenna complexes and the specific xanthophylls they bind play vital parts in tuning the function of the light-harvesting antenna to physiological need.

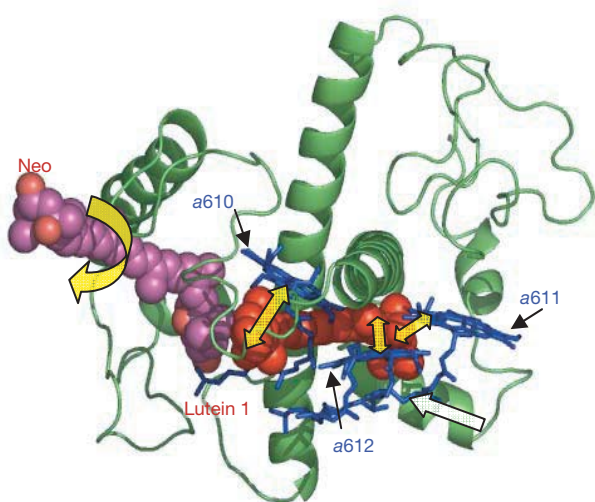


Figure 3 | Model illustrating the molecular mechanism of qE. Structural model of an LHCII monomer²¹ showing the key pigments involved in the establishment of qE. Lutein 1 (red) is closely associated with chlorin rings of chlorophyll *a* 610, 611 and 612 (blue, small black arrows). Curved broad yellow arrow, the configurational twist of the neoxanthin (Neo) molecule (pink); white broad arrow, the putative movement of lutein 1 towards the chlorophyll cluster (broad yellow arrows).

METHODS SUMMARY

LHCII trimers were prepared from spinach photosystem-II-enriched particles by isoelectric focusing, followed by sucrose gradient centrifugation²⁹. Oligomeric LHCII was obtained by incubation with SM-2 Absorbent (Bio-Rad). Chloroplasts from *Arabidopsis thaliana* were isolated and assayed as described previously³⁰. Room temperature fluorescence data (qE , K_d) are presented in the form $F(\text{unquenched}) - F(\text{quenched})/F(\text{quenched})$. Raman measurements were obtained at 77 K using 488 nm excitation^{10,14}, and the presented spectra normalized to 964 cm^{-1} unless stated otherwise. Transient absorption spectroscopy used 675 nm 100 fs laser pulses, with a repetition rate of 40 kHz and energy per pulse of about 10 nJ. The data were subjected to global and target analysis¹².

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 April; accepted 14 September 2007.

- van Grondelle, R. *et al.* Energy transfer and trapping in photosynthesis. *Biochim. Biophys. Acta* **1187**, 1–65 (1994).
- Horton, P., Ruban, A. V. & Walters, R. G. Regulation of light harvesting in green plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 655–684 (1996).
- Niyogi, K. K. Photoprotection revisited. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 333–359 (1999).
- Demmig-Adams, B. & Adams, W. W. III. Antioxidants in photosynthesis and human nutrition. *Science* **298**, 2149–2153 (2002).
- Horton, P. & Ruban, A. V. Molecular design of the photosystem II light harvesting antenna: photosynthesis and photoprotection. *J. Exp. Bot.* **56**, 365–373 (2005).
- Külheim, C., Ågren, J. & Jansson, S. Rapid regulation of light harvesting and plant fitness in the field. *Science* **297**, 91–93 (2002).
- Holt, N. E. *et al.* Carotenoid cation formation and the regulation of photosynthetic light harvesting. *Science* **307**, 433–436 (2005).
- Standfuss, J. *et al.* Mechanisms of photoprotection and nonphotochemical quenching in pea light-harvesting complex at 2.5 Å resolution. *EMBO J.* **24**, 919–928 (2005).
- Horton, P., Wentworth, M. & Ruban, A. V. Control of the light harvesting function of chloroplast membranes: the LHCII-aggregation model for non-photochemical quenching. *FEBS Lett.* **579**, 4201–4206 (2005).
- Pascal, A. A. *et al.* Molecular basis of photoprotection and control of photosynthetic light-harvesting. *Nature* **436**, 134–137 (2005).
- Robert, B. *et al.* Insights into the molecular dynamics of the plant light harvesting proteins *in vivo*. *Trends Plant Sci.* **9**, 385–390 (2004).
- van Stokkum, I. H. M., Larsen, D. S. & van Grondelle, R. Global and target analysis of time-resolved spectra. *Biochim. Biophys. Acta* **1657**, 82–104 (2004).
- Bassi, R., Pineau, B., Dainese, P. & Marquardt, J. Carotenoid-binding proteins of photosystem II. *Eur. J. Biochem.* **212**, 297–303 (1993).
- Ruban, A. V., Pascal, A. A. & Robert, B. Xanthophylls of the major photosynthetic light-harvesting complex of plants: identification, conformation and dynamics. *FEBS Lett.* **477**, 181–185 (2000).
- Li, X. P. *et al.* A pigment-binding protein essential for regulation of photosynthetic light harvesting. *Nature* **403**, 391–395 (2000).
- Li, X. P. *et al.* PsbS-dependent enhancement of feedback de-excitation protects photosystem II from photoinhibition. *Proc. Natl Acad. Sci. USA* **99**, 15222–15227 (2002).
- Frank, H. A. *et al.* Mechanism of nonphotochemical quenching in green plants: Energies of the lowest excited singlet states of violaxanthin and zeaxanthin. *Biochemistry* **39**, 2831–2837 (2000).
- Ma, Y. Z. *et al.* Evidence for direct carotenoid involvement in the regulation of photosynthetic light harvesting. *Proc. Natl Acad. Sci. USA* **100**, 4377–4382 (2003).
- Berera, R. *et al.* A simple artificial light-harvesting dyad as a model for excess energy dissipation in oxygenic photosynthesis. *Proc. Natl Acad. Sci. USA* **103**, 5343–5348 (2006).
- Lampoura, S. S. *et al.* Aggregation of LHCII leads to a redistribution of the triplets over the central xanthophylls in LHCII. *Biochemistry* **41**, 9139–9144 (2002).
- Liu, Z. *et al.* Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution. *Nature* **428**, 287–292 (2004).
- van Grondelle, R. & Novoderezhkin, V. I. Energy transfer in photosynthesis: experimental insights and quantitative models. *Phys. Chem. Chem. Phys.* **8**, 793–807 (2006).
- Moya, I. *et al.* Time-resolved fluorescence analysis of the photosystem II antenna proteins in detergent micelles and liposomes. *Biochemistry* **40**, 12552–12561 (2001).
- Morosinotto, T., Baronio, R. & Bassi, R. Dynamics of chromophore binding to Lhc proteins *in vivo* and *in vitro* during operation of the xanthophyll cycle. *J. Biol. Chem.* **277**, 36913–36920 (2002).
- Wentworth, M., Ruban, A. V. & Horton, P. Thermodynamic investigation into the mechanism of the chlorophyll fluorescence quenching in isolated photosystem II light harvesting complexes. *J. Biol. Chem.* **278**, 21845–21850 (2003).
- Yan, H. *et al.* Two lutein molecules in LHCII have different configurations and functions: insights into the molecular mechanism of thermal dissipation in plants. *Biochem. Biophys. Res. Commun.* **355**, 457–463 (2007).
- Demmig-Adams, B. & Adams, W. W. III. Photoprotection in an ecological context: the remarkable complexity of thermal energy dissipation. *New Phytol.* **172**, 11–21 (2006).
- Tang, Y. *et al.* Heat stress induces aggregation of the light harvesting complex of photosystem II in spinach plants. *Plant Physiol.* **143**, 629–638 (2007).
- Ruban, A. V. *et al.* Determination of the stoichiometry and strength of binding of xanthophylls to the photosystem II light harvesting complexes. *J. Biol. Chem.* **274**, 10458–10465 (1999).
- Crouchman, S., Ruban, A. V. & Horton, P. PsbS enhances nonphotochemical fluorescence quenching in the absence of zeaxanthin. *FEBS Lett.* **580**, 2053–2058 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from: UK Biotechnology and Biological Sciences Research Council (P.H., A.V.R.); the Netherlands Organization for Scientific Research via the Foundation of Earth and Life Sciences (R.v.G., H.v.A., J.T.M.K., R.B.) and a VIDI Fellowship (J.T.M.K.); Laserlab Europe; ANR (program caroprotect) (A.A.P., B.R.); and the INTRO2 EU FP6 Marie Curie Research Training Network. We thank K. K. Niyogi for the gift of seeds of the L17 *Arabidopsis* line.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.H. (p.horton@sheffield.ac.uk), B.R. (bruno.robert@cea.fr) or R.v.G. (rienk@few.vu.nl).

METHODS

LHCII preparation and analysis. LHCII trimers prepared from spinach photosystem II particles were incubated at a chlorophyll concentration of 1 mg ml^{-1} in a 2 ml volume in a $1 \text{ cm} \times 1 \text{ cm}$ cuvette in the presence of 0.01% *n*-dodecyl β -D-maltoside. Biobeads (300 mg; SM-2 Absorbent) were added. Chlorophyll fluorescence was monitored with a Walz MiniPam fluorimeter. The decline in fluorescence started after approximately 10 min and quenching reached a K_d of around 8–10 after 40 min. For Raman experiments, samples were taken at appropriate K_d values and immediately frozen in liquid nitrogen in the Raman sample holders. For transient absorption experiments the concentration of LHCII was adjusted to give $A_{670 \text{ nm}}$ 0.3–0.5 at the chlorophyll Qy maximum.

Chloroplasts and leaves. Chloroplasts were prepared from dark-adapted *Arabidopsis* plants. Chloroplasts were ruptured by brief osmotic shock in 5 mM MgCl_2 and measured in a reaction medium containing 0.33 M sorbitol, 20 mM HEPES, pH 7.8, 5 mM MgCl_2 and 0.1 mM methyl viologen. The chlorophyll concentration was $20 \mu\text{M}$. Chlorophyll fluorescence analysis was carried out using a Walz PAM101 fluorimeter. A light saturation pulse was given at the ends of the light and dark periods to give the F_m' and F_m levels respectively. qE was calculated as $F_m - F_m' / F_m'$. Chloroplasts were measured in a $1 \text{ cm} \times 1 \text{ cm}$ cuvette. Samples taken after 5 min illumination ($1,500 \mu\text{mol m}^{-2} \text{ s}^{-1}$ light) and following 5 min dark relaxation were transferred to the Raman sample holder and immediately frozen with liquid nitrogen. Detached, dark-adapted *Arabidopsis* leaves were placed in the Raman sample holder and chlorophyll fluorescence analysis was carried out using the same protocol as for chloroplasts, and similarly frozen, immediately after the light or dark treatment.

Pigment analysis. Pigment compositions of LHCII and chloroplasts were determined by HPLC using a reverse phase C18 column (Merck LiChrospher 100 RP-18) and Dionex chromatography system. The solvent system was (solvent A, 87% acetonitrile, 10% methanol, 3% 0.1 M TRIS, pH 8; solvent B, 80% methanol, 20% hexane). The gradient from solvent A to solvent B was run from 9 to 12.5 min at a flow rate of 1 ml min^{-1} . Each peak was integrated at its optimum absorbance and analysed using Dionex Chromeleon software. The analysis showed that the LHCII samples contained only trace amounts of violaxanthin ($0.3 \pm 0.2\%$ of the total carotenoid) and no detectable antheraxanthin or zeaxanthin.

Transient absorption spectroscopy. Transient absorption spectroscopy used femtosecond pulses obtained from a titanium:sapphire oscillator-regenerative amplifier (coherent MIRA seed and RegA). The repetition rate was 40 kHz and the initial pulse $\sim 60 \text{ fs}$ at 800 nm. The beam was split into two, one of which (the probe beam) was focused on a CaF_2 plate to generate a white-light continuum. The other beam was used to pump an optical parametric amplifier to obtain the pump beam at 675 nm ($\sim 100 \text{ fs}$). The energy per pulse was about 10 nJ. The polarization between the pump and probe beams was set at the magic angle. High repetition rate, single-shot multi-channel detection and rejection of outliers improved the signal-to-noise ratio by about a factor of ten (compared to a standard pump-probe experiment).

naturejobs

JOBS OF THE WEEK

Several months ago, I noticed a trend in the stream of e-mails coming into the Naturejobs inbox. Scientists and students were writing to our “Ask an Expert” column (<http://tinyurl.com/3748bo>) seeking advice about careers in biotechnology. This, of course, isn’t unusual. But nearly all the e-mailers were from India and most had completed master’s degrees in biotechnology there. It didn’t take much sleuthing to discern a trend worth pursuing. The result is this week’s feature (see page 580).

India’s biotech industry is booming. At least, that’s what the media reports say. So one can understand Indian scientists being optimistic about their job prospects in the field. Unfortunately, the sector is not yet generating enough jobs to cope with the increasing number of graduates.

So why is there such a discrepancy between the hype and the reality? At the moment, most of the jobs in the sector are in biotech services, such as outsourcing activities related to clinical trials. Few opportunities are available in innovation and discovery — the type of jobs that graduates are looking for and that would drive the field forward. In addition, the country’s education system has not kept up with the sector’s growth. Many programmes lack sophistication in terms of equipment, practical experience offered and the number of qualified faculty members. Often, master’s graduates simply don’t have the necessary skills to enter the workforce.

The government has recognized the shortfall and is seeking ways to remedy the situation. Together with the Association of Biotechnology Led Enterprises, it is setting up ‘biotechnology finishing schools’ that would provide the practical experience needed by master’s biotech graduates. The move is similar to the development of ‘professional’ master’s degree programmes by many US universities (see *Nature* **445**, 458; 2007).

Young researchers-in-training must be careful not to believe the media hype. And, perhaps most importantly, they should make sure that the degree programme they choose provides them with the training they need to further their career.

Gene Russo, acting editor of Naturejobs

CONTACTS

Acting Editor: Gene Russo

European Head Office, London

The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:

Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com

Natureevents:

Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:

Muriel Lestringuez (4994)

Southwest UK/RoW:

Nils Moeller (4953)

Scandinavia/Spain/Portugal/Italy:

Evelina Rubio-Hakansson (4973)

Northeast UK/Ireland:

Matthew Ward (+44 (0) 20 7014 4059)

North Germany/The Netherlands:

Reya Silao (4970)

South Germany/Austria:

Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:

Stephen Russell
To send materials use London
address above.
Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Jasmine Myer

US Head Office, New York

75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:

Ayako Watanabe
Tel: +81-3-3267-8765
e-mail: a.watanabe@natureasia.com



Growth factor: Bangalore is one of India's biotech hubs.

Indian biotech's bumpy road

Vinay Vasant didn't think twice about returning to India from England, after a year-long biotechnology master's at Newcastle University. He missed home and had heard a lot about India's blooming biotechnology sector. The career prospects, he predicted, would be bright. Although most of his Indian classmates stayed on in Britain, Vasant packed his bags and flew to Pune.

A year later, Vasant is wondering whether he should have boarded the plane. Despite his foreign credentials and several promising interviews, Vasant has yet to land a job. "I thought I would come back and get opportunities, but I am sitting at home," he says. "A lot of young students now ask me for advice on whether they should go into biotech, but I don't know whether to encourage or discourage them."

Vasant's confusion reflects a growing anxiety among graduates who are struggling to find a foothold in India's growing biotechnology market. It's been a rude awakening for those who flocked to the field after widespread media reports likened the growth of Indian biotech to the country's startling information-technology boom that began in the 1990s. Cashing in on that perception, hundreds of private institutes sprang up and now churn out thousands of bachelor's- and master's-level biotechnology graduates every year.

Unfortunately, the hype has outstripped reality, say industry observers. India's biotech industry is growing, but hasn't matured enough to absorb all the country's fresh talent. Making matters worse, the quality of biotech education is often inadequate. Despite the many degree holders, industry officials say it has been a chore recruiting talent with the requisite skills. "Unless we do something now, human resources will be a big limitation for the growth of India's biotechnology

The rush to join in India's latest boom sector has led to a bottle-neck, says Paroma Basu.



Hands on: a student on the master's course in Baroda.

industry," says Kottaram Narayanan, president of the Association of Biotechnology Led Enterprises (ABLE) and managing director of agri-biotech company Metahelix Life Sciences in Bangalore.

Indian biotech generated revenues of some \$2 billion in 2006–07, 60% of it from exports, according to the latest industry survey by ABLE and trade publication *BioSpectrum India*. Biopharmaceuticals — particularly the development of generic drugs and affordable vaccines — is the sector's biggest profit-maker, generating about 70% of total biotech revenues fuelled, in part, by profitable partnerships and mergers with companies in the West and other parts of Asia. The 'bioservices' market — typically drug companies outsourcing clinical-trial tasks to contract research organizations (CROs) — takes in around 13% of the total biotech pie. *Bt* cotton, the only transgenic agricultural product on the market, has made about 10%. Biomanufacturing and bioinformatics are also starting to play a part.

Innovation needed

Yet India's biotech market is still finding its feet when it comes to home-grown biotech inventions and discoveries. For a real boom, Indian scientists need to start innovating, says Virender Kumar Vinayak, president of biopharmaceutical R&D at health-care company Panacea Biotec in New Delhi. The emphasis should be on developing new biomolecules, tools and medical devices. This will require a steady supply of critically thinking scientists with solid hands-on skills. But therein lies the problem.

The inconsistent and largely unregulated biotech education sector is producing up to 30,000 graduates every year, according to reports from *BioSpectrum*

India. “A large number of teaching shops have opened up in India, but most of these churn out improperly trained biotechnologists,” says Vinayak. There is a dire need to improve course curricula to meet the requirements of the industry, he adds.

Many of the estimated 300 private programmes lack lab space and basic equipment. “Some of these graduates have never even seen a gel apparatus,” says Rajeev Soni, chief operating officer of CRO Premas Biotech, based near Delhi. Soni says Premas has to train graduates for at least six months before they are ready for the job. “This lag time is really hurting the industry,” he says.

Krishna Ella, managing director of vaccine-maker Bharat Biotech in Hyderabad, says that his company receives 300–500 applications for every new job opening, but the vast majority of candidates don’t fit the bill. “It’s not just a degree that is important,” Ella says. “The most important things are practical skills and the ability to think critically as a scientist.” Worsening matters, there’s a dearth of good teachers at private programmes, particularly in small or remote towns.

Aparajita Mitra, who did a bachelor’s degree in biotechnology at Fergusson College in Pune, laments the lack of hands-on training and equipment. Dissatisfied with her career prospects in biotech, she is pursuing an MBA. Barely one-third of her college class of 30 are still working in biotech, with most stuck in underpaid administrative and technician positions, Mitra says.

Not all private programmes are poorly equipped. New Delhi’s Amity University and Tamil Nadu’s SRM University, for example, have spacious labs and cutting-edge equipment such as protein-purification and electrophoresis systems. These programmes have the money to attract qualified staff. SRM also has ties with industry, enabling bachelor’s students to do company internships in their final year. But such schools charge up to 100 times more — around US\$4,000 in tuition fees per year — than government programmes.

India’s government has actually fostered some strong biotech feeder programmes. In the mid-1980s, prime minister Rajiv Gandhi gave \$500,000 to five universities to establish world-class biotech degrees; they bought modern research tools and recruited high-profile Indian scientists from around the world. The Department of Biotechnology (DBT) funds 60 programmes producing some 1,000 graduates every year, according to Suman Govil, a DBT adviser for human-resource development. The government awards recurring grants of up to \$75,000 a year to each of its sponsored programmes, and



“The most important things are practical skills and the ability to think critically as a scientist.”
— Krishna Ella

has spent more than \$5 million this year on biotech teaching and training alone, Govil says.

“Our students have not really faced any problems with regard to finding opportunities,” says Rakesh Bhatnagar, chair of the biotechnology centre at Delhi’s Jawaharlal Nehru University, who left a research career at the US National Institutes of Health to come to Delhi. The university was among the first five to receive the government’s biotech grant and is regarded as one of India’s top biotech centres. Only 20–30 students get in, by way of a competitive exam. They learn basic sciences such as immunology and genetics in the first year and spend another year working on a real research problem in the lab, says Bhatnagar.

Expansion limited

But even India’s best institutes have struggled to recruit experienced staff, says geneticist Bharat Chattoo, coordinator of the biotech programme at the Maharaja Sayajirao University of Baroda, one of India’s first feeder programmes. The government hasn’t helped much, he says: “Most places have not got additional faculty positions after their programmes were started. How is the programme meant to rejuvenate and grow?”

Even qualified graduates are finding few jobs with a real biotech focus, says Kumaraswamy Ramasamy, dean of SRM’s school of biotechnology. Every year, almost half of SRM’s 180 BSc graduates end up going into the IT sector to work in bioinformatics companies, Ramasamy says. Many others find work at CROs. To expand skill sets and attract an array of employers, SRM is giving mandatory computer courses and emphasizing communication and languages. This flexibility sets private schools apart from government programmes, which accept far fewer students and focus entirely on research, Ramasamy adds.

Although that approach may serve a short-term need in the market, Chattoo believes that it’s not seeding a new crop of scientific innovators. “We need to provide an ecosystem in which innovations can thrive,” he says. “My role as a teacher is not to produce people who are vocationally trained but people who can be leaders.”

Government and industry officials are now kick-starting various measures to ease the mismatch between academia and industry. The DBT offers short-term training courses for biotechnology teachers around India. The department also runs a programme that places fresh master’s graduates in six-month, paid industry internships. It placed 200 students last year, and intends to place at least 500 this year, Govil says.

Meanwhile, ABLE, in collaboration with a government advisory group, plans to start a series of competitive ‘biotechnology finishing schools’ for master’s-level graduates who would spend six months to a year working on a real industry project while honing fundamental science concepts and analytical skills. The first such venture — with an initial group of 100 students — is due to start with the 2008 academic year in the southern state of Karnataka. Narayanan and others hope the model, if successful, will be copied in other parts of the country. “Once a person in biotechnology is properly trained, he or she will be immediately picked up by companies,” says Narayanan. But this runs counter to the experiences of young graduates such as Vasant, who are anxiously waiting for the Indian biotech sector to live up to the hype. ■

Paroma Basu is a freelance writer in New Delhi, India.



Branching out: students on the biotech programme in Baroda are learning to be leaders.

MOVERS

**Susan Avery, president and director,
Woods Hole Oceanographic Institution,
Woods Hole, Massachusetts**



2004–05 and 2006–07:

Interim vice-chancellor for research and dean of the graduate school, University of Colorado, Boulder

2005–06: Interim provost and executive vice-chancellor for academic affairs, University of Colorado, Boulder

1994–2004: Director, CIRES, University of Colorado, Boulder

Susan Avery's PhD adviser Marvin Gellar saw her potential right away — and not just in terms of scientific acumen. "I predicted she would go further than her considerable scientific talents alone might merit because of her excellent people skills and effective collaborations," says Gellar. With Avery due to start as president of the Woods Hole Oceanographic Institution, Gellar's prediction looks sound.

Avery began her undergraduate study in physics at Michigan State University in East Lansing, but soon decided to focus on the physics of the natural world. She considered oceanography, then opted to study how atmospheric waves propagate in the stratosphere as a PhD student at the University of Illinois at Urbana-Champaign. Upper-atmospheric physics research programmes are often found in electrical-engineering departments because the region, dubbed the ionosphere, controls how radiowaves move through the atmosphere. This engineering experience opened the door to Avery's first faculty position, in the University of Illinois' electrical-engineering department.

When her husband got a job in Boulder, Avery amassed two fellowships — from the National Science Foundation and the University of Colorado's Cooperative Institute for Research in Environmental Science (CIRES) — to start her own research. She quickly built collaborations at the National Center for Atmospheric Research and the National Oceanic and Atmospheric Administration, and soon had a faculty position at the University of Colorado at Boulder.

Avery then developed new radar technology that allowed the first measurements of wind in the upper atmosphere of remote equatorial regions, leading to years of collaborations in the tropics. "My work began to flourish by bringing point observations into the global context," she says.

After earning tenure, she reluctantly accepted a position as associate dean of research and graduate education. Before long, she became director of CIRES, where she spent a decade creating interdisciplinary programmes, and realized she enjoyed that part of the job.

Known as a strong administrative leader, Avery has held several key interim positions at the university since it underwent drastic changes between 2005 and 2007. Then she sought a change, and found it at the Woods Hole in Massachusetts. "I admit it was scary to accept a position leading a different field of scientists," she says. Among her initial goals at Woods Hole will be addressing the ocean's role in climate change.

Virginia Gewin

BRICKS & MORTAR

Grand designs

Ceiling patterns representing cell division and a colour scheme inspired by chromosome painting are among the decorative features of the new Cancer Institute at University College London (UCL). The £40-million (US\$84-million) building was partially funded by the childhood cancer charity Children with Leukaemia, the Wolfson Foundation and Atlantic Philanthropies.

The institute nestles between the new University College Hospital and the Wolfson Institute for Biomedical Research, with which it will share several core services, including a microarray facility, a proteomics unit, and imaging and transgenesis suites.

The institute will accommodate 300 scientists, doubling the number working on cancer at UCL. With excellent clinical resources nearby, including the National Hospital for Neurology and Neurosurgery, tumour banks and one of the largest bone-marrow transplant programmes in Europe, the UCL Cancer Institute hopes to boost basic and translational research in the area.

Director Chris Boshoff and his colleagues aim to investigate haematological malignancies, brain cancer, adolescent and young-adult cancers, and head, neck and lung cancers. "We will focus on cancer

types that aren't always at the top of other people's priorities," says Boshoff.

Over the next two years, Boshoff will recruit up to six promising early-career scientists to their first group-leader position. "We can help them establish their laboratory and then they can start their careers with us," he says.

Among the 180 scientists already there, Boshoff has secured some top talent — including leaders in genomics, brain surgery and paediatric and adolescent cancer biology — from other institutions, such as the Wellcome Trust Sanger Institute in Cambridge, the Hospital for Sick Children in Toronto and the University of California, San Francisco.

Canadian recruit Poul Sorensen, a clinician-scientist and expert in childhood cancers, says that the funding and the leadership helped lure him overseas.

"A lot of people talk about translational research, but my feeling is the people at the UCL Cancer Institute really understand what it means," he says. "Those driving the institute are clinician-scientists with a strong background in basic science. You need that combination to have the vision to develop new strategies for cancer treatment."

Hannah Hoag

POSTDOC JOURNAL

Leaving science research

A couple of weeks ago, and with much trepidation, I knocked on the door of my adviser's office. I was about to tell him of my plan to leave the world of basic science research, and I had no idea how he would react. I had tried to imagine all of the possible outcomes of our conversation, and although I suspected that I could deal with any consequences, I honestly wasn't sure what would happen. Nevertheless, my growing discontent with lab research made it a conversation that I could no longer put off.

To my delight, my adviser was very understanding of my desire to pursue other interests, and immediately asked what I wanted to do and how he might be able to help. As a result of our conversation I have decided to continue working in the lab for now, focusing my attention less on the basic biological processes and more on the multiple public-policy issues related to our area of research. Given my burgeoning interest in policy matters, this is definitely a step in the right direction.

In my first journal entry, I gave myself two to three years to determine whether I wanted to make a career out of science research. A year later, I've concluded that it's time to start exploring other options. After all, there's much out there to explore, and there's no time to spare.

Peter Jordan is a visiting fellow at the National Institute of Diabetes and Digestive and Kidney Diseases in Bethesda, Maryland.

Dating for the wired generation

A match made in silicon.

Stephen Gaskell

First there was Gloria. I met her through a dating agency called Heaven Scent. They were big on smells. You know when you sniff your lover's hair and get all soppy and light-headed? Do you think it's because there's something innately appealing about their particular blend of Head and Shoulders and sweat?

Didn't think you did.

What's really going on is that you're testing your and your partner's disease-defence compatibility. Apparently the make-up of your personal aroma (or 'funk' as is the case) is a pointer to the subtleties of your immunological system. Get a bad combination and your darling cherubs are going to grow up as sickly little mites not even able to stave off the common cold. And who wants a snivelling brat twenty-four seven?

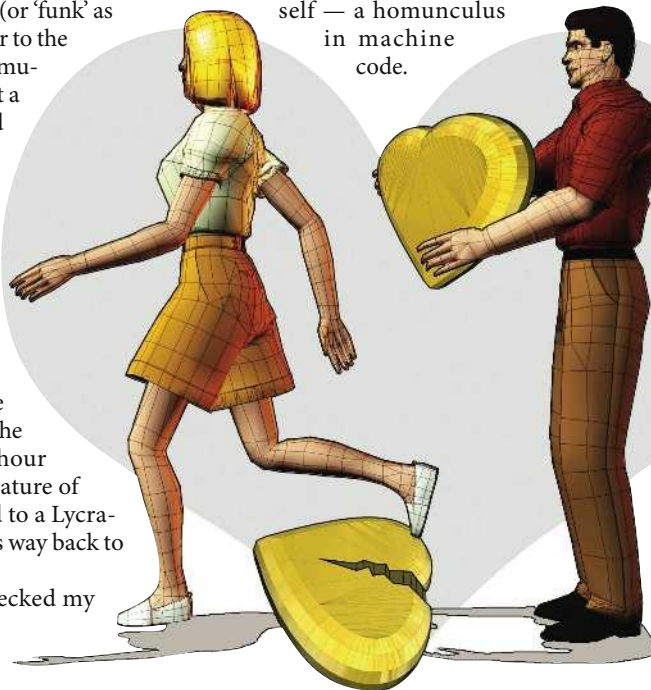
So, anyway, the testing kit arrived at the office, and half-an-hour later a chemical signature of yours truly, strapped to a Lycra-clad courier, made its way back to Heaven Scent's labs.

The next day I checked my inbox and, lo and behold, they'd found me a match! They even attached her profile and a pic. Gloria was smart and cute — as well of great olfactory promise. That evening, I showered and shaved — taking care not to hide my approved scent under a swathe of deodorant — and headed out for dinner. As soon as we met I knew she wasn't the one. She was a Chanel belle — drenched in the most nauseous perfume imaginable. And she had a caustic personality to match. Heaven sent? Dream on!

Next came Lucy. Lucy was on Perfect Balance's books. They're a dating website who place their faith in facial symmetry. They reckon that we're most attracted to those whose physiognomy matches our own. Being a man whose face looks like a grizzly bear's, I had my doubts. I did the mug-shots, anyhow — it only took five minutes in front of my workstation's webcam.

When Lucy approached my table at Starbucks, my first thought was that she was a guy looking for directions to the gents. Then I did a double-take and thought I was looking in the mirror. I know we're told to love ourselves, but that was taking the mantra too far. A double-mocha latte later, and I was outta there.

Perseverance, I said to myself, is the key. I went to Avatar Matchmakers — virtually. They've got offices in Sim-London. Here's the deal: you make a personality profile, give your physical stats, and then they create an online version of yourself — a homunculus in machine code.



This virtual-you then breezes around Sim-London trying to find love with the other virtual bods. When it finds a match, the real you in the real world goes and meets the man, or woman, behind the ones and noughts. It's a genius idea. Minimal effort, maximum encounters. You can even watch your future date's virtual-self going about his or her business just to make sure they're your type. The realism of these avatars is incredible. When virtual-me found virtual-Sally I thought I'd found the one.

Boy, was I mistaken. Virtual-Sally and real-Sally couldn't have been more different. Virtual-Sally goofed around, cracked jokes, and had a lot of class. Real-Sally moved like a marionette, jabbered on and on, and couldn't have been more rude to the barista who served us.

She said I didn't live up to her expectations either.

I guess we're our own worst judges. The date ended badly. I'm just glad she was drinking iced — rather than hot — coffee when she threw her drink over me.

Still, I couldn't stop myself. Might as well try them all, I thought. Vocal harmony matching struck a chord with Beverly. Traumatic life-event compatibility threw me in with Kate. Biorhythm affinity coaxed me into a date with Pritha. All were disasters.

Next I tried the Wicked Ways Dating Emporium. They matched partners through their MQ — their Morality Quotient. The theory being that we're attracted to those who are just as bad — or good — as ourselves. In a historical free-for-all, Jesus would've got paired with Mother Theresa, whereas Hitler would've smooched with, well, Eva Braun, I guess.

My MQ was 95. A little less virtuous than average. Bad enough to hit on my ex's cousin. Not bad enough to hit on my ex's sister. I got Jasmin.

"So you're the bad boy," she said, looking me up and down as we greeted each other outside the Casa Blanca Tapas Bar.

"And you're the bad girl!" I grinned. Mis-

chief was this girl's middle name. Or so I thought.

At dinner we ordered nearly every dish on the menu, and stuffed ourselves silly. Several bottles of vintage Rioja helped to loosen our tongues. It was a great date — until it was time to leave.

Jasmin leaned close and eyed the door, "Ever skipped on the bill?"

Before I could answer she was halfway across the restaurant, napkins scattering in her wake. The girl was evil. I couldn't run on a dinner bill. I slumped back in my chair, defeated.

That was the last straw. Bring on the Luddites. I'm going to a bar.

Stephen Gaskell is currently taking an MSc in evolutionary and adaptive systems at the University of Sussex, UK. He recently sold his first story to *Writers of the Future* Volume XXIII.

JACEY